Future Increases in North American Extreme Precipitation in CMIP6 Downscaled with LOCA

DAVID W. PIERCE,[a] DANIEL R. CAYAN,[a] DANIEL R. FELDMAN,[b] AND MARK D. RISSER[b]

[a] *Division of Climate, Atmospheric Sciences, and Physical Oceanography, Scripps Institution of Oceanography, La Jolla, California*
[b] *Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, California*

(Manuscript received 1 November 2022, in final form 27 January 2023, accepted 7 March 2023)

ABSTRACT: A new set of CMIP6 data downscaled using the localized constructed analogs (LOCA) statistical method has been produced, covering central Mexico through southern Canada at 6-km resolution. Output from 27 CMIP6 Earth system models is included, with up to 10 ensemble members per model and 3 SSPs (245, 370, and 585). Improvements from the previous CMIP5 downscaled data result in higher daily precipitation extremes, which have significant societal and economic implications. The improvements are accomplished by using a precipitation training dataset that better represents daily extremes and by implementing an ensemble bias correction that allows a more realistic representation of extreme high daily precipitation values in models with numerous ensemble members. Over southern Canada and the CONUS exclusive of Arizona (AZ) and New Mexico (NM), seasonal increases in daily precipitation extremes are largest in winter (~25% in SSP370). Over Mexico, AZ, and NM, seasonal increases are largest in autumn (~15%). Summer is the outlier season, with low model agreement except in New England and little changes in 5-yr return values, but substantial increases in the CONUS and Canada in the 500-yr return value. One-in-100-yr historical daily precipitation events become substantially more frequent in the future, as often as once in 30–40 years in the southeastern United States and Pacific Northwest by the end of the century under SSP 370. Impacts of the higher precipitation extremes in the LOCA version 2 downscaled CMIP6 product relative to the LOCA downscaled CMIP5 product, even for similar anthropogenic emissions, may need to be considered by end-users.

KEYWORDS: Extreme events; Precipitation; Downscaling; Climate models; General circulation models; Regional models

1. Introduction

Extreme precipitation has serious societal and economic impacts, potentially triggering floods that can cause property damage or loss of life. For instance, between 2008 and 2021 the United States experienced 22 flooding events that each caused over $1 billion in damage (adjusted for inflation; NCEI 2022). Recognizing the potentially disastrous effects of floods, it is common to build water management infrastructure that includes a goal of minimizing flooding, with designs informed by historically observed precipitation and streamflow extremes (NRC 1982). However, this approach becomes problematic in a world with a changing climate, where historical extremes may not indicate the full range of future events (e.g., Milly et al. 2008; Arnell and Gosling 2016; Donat et al. 2016).

A tendency toward increasing precipitation extremes is likely in a warming world, since the saturation vapor pressure of water vapor increases with temperature, leading to greater atmospheric burdens of water vapor to fuel precipitation events (the so-called thermodynamic effect; e.g., Westra et al. 2014; Kröner et al. 2017; Norris et al. 2019; Harp and Horton 2022). This response can be regionally modulated by changes in atmospheric circulation (the dynamic effect), which can alter the trajectory of water vapor transport, the atmosphere's vertical stability, and moisture availability (e.g., Martinkova and Kysely 2020). The poleward shift of storm tracks is one example of anticipated atmospheric circulation change in coming decades (e.g., Yin 2005). As another example, in the western United States, atmospheric rivers (ARs) carry an enormous water vapor flux from the Pacific Ocean onto land, fueling extreme precipitation events that can generate flooding with extensive damages that increase exponentially with flood magnitude (Corringham et al. 2019). A string of ARs caused the costly Oroville Dam spillway failure in 2017 (Vano et al. 2019). Future changes in ARs, including their water vapor burden, landfalling latitude, and orientation with respect to topographical features will affect extreme precipitation in the region (Gershunov et al. 2019; Tan et al. 2020; Michaelis et al. 2022). A review of various hydroclimate changes due to projected changes in atmospheric circulation is given in Zappa (2019), including changes in the Mediterranean region of Europe, the eastern North Pacific, and the southern part of Chile. Other evaluations of projected atmospheric circulation changes can be found for Europe (Ozturk et al. 2022), China (Yang et al. 2021), the polar regions (Screen et al. 2018), and New Zealand (Sturman and Quenol 2013), among other locations.

Some of the best tools to examine future changes in precipitation extremes are Earth system models (ESMs). The output from coordinated experiments using the latest generation of ESMs is collected in the Climate Model Intercomparison Project version 6 (CMIP6) archive (Eyring et al. 2016). CMIP6 holds data from numerous ESMs constructed by researchers across the globe, incorporating varied scenarios of future

Corresponding author: David Pierce, dpierce@ucsd.edu

DOI: 10.1175/JHM-D-22-0194.1

greenhouse gas emissions, anthropogenic aerosol emissions, and land use changes (O'Neill et al. 2014, 2016; Riahi et al. 2017). Many previous studies have examined extreme precipitation in the CMIP6 models. For example, Srivastava et al. (2020) and Akinsanola et al. (2020) examined model skill relative to observations over the CONUS; Ge et al. (2021) over Southeast Asia; Zhu et al. (2020), Xu et al. (2022), and Luo et al. (2022) over China; Gupta et al. (2020) over India; Chen et al. (2021) over the western North Pacific and East Asia; Dike et al. (2022) over central Asia; Faye and Akinsanola (2022) over West Africa; and Kim et al. (2020) over the entire world. Many studies note that the CMIP6 models are somewhat more realistic than the CMIP5 models, a result that Pierce et al. (2022), examining the upper Colorado River basin, found to be due to improvements in the depiction of atmospheric circulation while systematic biases remained little changed from the CMIP5 models. Additionally, these studies generally noted that the multimodel ensemble average is typically better than individual models, which has been shown to arise from the way errors cancel across different models (e.g., Pierce et al. 2009).

Although ESMs are a key source of information for understanding future changes in extreme precipitation, they have two significant drawbacks when used to anticipate and mitigate impacts of climate change on a regional or local scale. First, the coarse grid spacing of most ESMs is not sufficient to fully resolve the influence of topography on precipitation in mountainous regions. Although some CMIP6 ESMs have 50-km spatial resolution, most are on the order of 100 km with some up to 250 km. At these resolutions the effect of topography on precipitation is muted, since the topography itself is poorly represented. In addition, coarser resolution models have a harder time capturing atmospheric processes that drive precipitation extremes (e.g., Li et al. 2011; O'Brien et al. 2016). Second, CMIP6 ESMs have biases in their output, including in the depiction of extreme precipitation, that arise from model errors. Some stakeholders require realistic, unbiased time series of daily precipitation values for input to existing application or impact models. For these stakeholders, ESM biases can render an existing application model unusable, especially if the application model has a nonlinear dependency on the input variables. For example, land surface models typically fall in this category, since runoff depends in a nonlinear way on the magnitude and temporal characteristics of the precipitation time series. Another example is utility-scale electricity demand, which is a nonlinear function of air temperature. For such users, removing biases in the ESM data is a requirement for the application models to be useful.

The need for more finely resolved, unbiased climate projections is typically addressed by bias correcting and spatially downscaling the original ESM data, although such techniques cannot obviate inherent model deficiencies (e.g., Maraun et al. 2017). There are numerous approaches for accomplishing this. For example, bias correction over the historical period can use quantile mapping (QM; e.g., Panofsky and Brier 1968; Thrasher et al. 2012; Maraun 2013), although QM is not appropriate for bias correcting future climate projections since it distorts future climate change in a nonphysical way (Hagemann et al. 2011; Pierce et al. 2013b; Maurer and Pierce 2014). Approaches for bias correcting future climate projections include univariate methods such as cumulative distribution function transform (CDFt) (Michelangeli et al. 2009) and equidistant cumulative distribution function matching (EDCDFm) (Li et al. 2010); conditional multivariate methods such as PresRat (Pierce et al. 2015), which was used in localized constructed analog (LOCA) version 1, and fully multivariate approaches such as Cannon (2018) and Guo et al. (2019). Multivariate approaches try to preserve between-variable relationships that are important to climate impacts, such as the temperature when precipitation occurs, which determines winter snow cover. Reviews of bias correction for climate and hydrological applications are given in Teutschbein and Seibert (2012) and Maraun (2016).

Spatial downscaling likewise encompasses a broad range of techniques, including dynamical methods that employ numerical models similar to weather prediction models and statistical approaches that use historically observed relationships between weather phenomena at large and small scales to translate coarse-resolution ESM fields into a more finely resolved representation (e.g., Wood et al. 2004; Fowler et al. 2007; Gudmundsson et al. 2012; Pierce et al. 2013b; Bukovsky et al. 2015; Jacob et al. 2020). Dynamical downscaling may do a better job describing climate in coming decades, when nonstationarity has made historically observed relationships a questionable guide for future climate changes. However, even dynamical models employ parameterizations that are partly developed from historical observations, and therefore may contain implicit stationarity assumptions that misrepresent future changes. Furthermore, dynamical downscaling output typically still has biases that need to be corrected for end user applications, which requires a statistical bias correction step with its own stationarity assumptions. Some dynamically downscaled products are still too coarse for climate adaptation needs [e.g., NARCCAP (Mearns et al. 2009) provides data at a resolution of 50 km] and therefore require further downscaling and bias correction for some applications (Ahmed et al. 2013). Additionally, dynamical downscaling is computationally expensive, which limits the total number of models, emissions scenarios, and ensemble members that can be downscaled, which limits the range of conditions that can be explored for end-users.

Statistical downscaling methods are more overt and fundamental in the stationarity assumptions included, which might misrepresent future climate changes to some degree. However, they have a much lower computational cost, so many more models, greenhouse gas emission scenarios, and ensemble members can be downscaled with a given computational budget. Since different ESMs project a wide variety of future climate changes, having enough models to both encompass the projected range and identify rare combinations of climate extremes for infrastructure stress tests are important considerations for many stakeholders.

Statistical downscaling encompasses a wide range of techniques, such as regression analysis between large-scale climate indices or patterns and a particular location (e.g., Dibike and Coulibaly 2005; Chen et al. 2014), neural networks and other

machine learning approaches (e.g., Sachindra et al. 2018; Sharifi et al. 2019), and weather typing, which relates large-scale circulatory regimes to weather at a point (e.g., Boé et al. 2007; Van Uytven et al. 2020). Besides the stationarity assumption noted previously, statistical techniques require high-quality observations, known as training data, at both the large scale and at the locations to be downscaled to in order to develop the statistical relationships used to estimate information at the finer spatial resolution. This can be an important limitation to statistical downscaling in some circumstances, such as when trying to downscale variables that are not routinely observed (e.g., hub-height wind speeds for wind turbine applications) or when a region has a low density of observing stations (e.g., in Alaska). The results of statistical methods depend on the training data used, and differences in training data can account for important differences between different statistically downscaled datasets (e.g., Wang et al. 2020; Vano et al. 2020). Some statistical methods use a different variable as predictor than as the predictand (variable to be predicted). For example, weather typing approaches might predict precipitation using atmospheric pressure and water vapor fields. LOCA belongs to the class of methods that use the same variable for the predictor and predictand, and is an analog day method, where historical days that are similar to the model day being downscaled are used as the basis of constructing the downscaled field. LOCA has produced good results downscaling temperature, precipitation, humidity, and the heat index in cross-validation tests and when compared to other similar statistical methods (Pierce et al. 2014, 2015; Pierce and Cayan 2016) and provided fundamental data for the Fourth National Climate Assessment (Avery et al. 2018) and California's Fourth Climate Change Assessment (Pierce et al. 2018). More details on LOCA's approach are given in section 2.

In this work we investigate future precipitation extremes generated by a new version of the LOCA (Pierce et al. 2014) statistical downscaling method. The downscaled output covers a domain from central Mexico through southern Canada at 6-km spatial resolution. The improved representation of daily precipitation extremes is accomplished through two changes: 1) using an improved precipitation training dataset that itself better represents daily extremes (Pierce et al. 2021); 2) adding an ensemble version of bias correction to the process (described below). The primary focus of this work is on point 2, the ensemble bias correction for precipitation. Additionally, the new version of the downscaled dataset, which we term LOCA version 2, is produced using the latest CMIP6 models rather than on CMIP5 models used in LOCA version 1. Our goal is to document the projected changes in extreme precipitation found in this new downscaled dataset for various time horizons and emission scenarios.

The LOCA version 2 output includes daily temperature minimum and maximum in addition to daily precipitation, but we focus on the latter here since the main changes from LOCA version 1 (CMIP5) are in the treatment of precipitation. Section 2 covers the changes in precipitation training data and the ensemble bias correction methodology and lists the models and ensemble members that have been downscaled. Section 3 shows the results in terms of projected changes in 5-, 50-, and 500-yr return values of daily precipitation as a function of season, location, and SSP, focusing on multimodel ensemble averages for brevity. A discussion and some considerations for users of the data are given in section 4, and conclusions in section 5.

2. Data and methods

a. *LOCA statistical downscaling*

Localized constructed analogs (Pierce et al. 2014) is a statistical technique for downscaling climate model output to more finely resolved spatial scales. LOCA requires as input a coarse-resolution ESM field to be downscaled and a fine-spatial-resolution set of training data that is generally obtained from observations. LOCA then produces a fine-spatial-resolution version of the original ESM field.

Analog downscaling methods such as LOCA work by finding a match between the ESM field being downscaled and the observations coarsened to the ESM grid. Conceptually, the best matching observed day is termed the analog day, and the original fine-spatial-resolution version of the analog day is the downscaled result. Because of the exceedingly high degrees of freedom for weather patterns, a single analog day rarely provides a satisfactory match to the ESM field being downscaled in any but the smallest domains (Van den Dool 1994). Constructed analog methods therefore combine multiple matching historical analog days into one overall "constructed" analog day to produce the fine-resolution downscaled result. Traditionally this construction is done via a weighted average of the best-matching analog days over the entire domain (e.g., Van den Dool 1994; Maurer et al. 2010; Abatzoglou and Brown 2012). However, the weighted average approach leads to a reduction in extremes for all variables and an increase in drizzle days when downscaling precipitation (Pierce et al. 2014). Additionally, this approach does not scale well with domain size, since a sufficiently large domain (e.g., the CONUS) includes locations separated by enough distance to have weather that is uncorrelated, making it harder to find observed days that happen to match the model day in multiple unrelated locations.

To address these problems LOCA combines the analog days spatially (Pierce et al. 2014), roughly analogous to assembling a jigsaw puzzle. The analog day selection proceeds in two steps. First, a pool of 30 best-matching analog days is chosen on a larger regional synoptic scale. Second, the best single matching analog day of the 30 is chosen from a smaller local region around the grid cell being downscaled to. This multiscale matching approach avoids averaging multiple analog days and so better preserves extremes, a key driver of climate change impacts, and eliminates the problem of drizzle days. Additionally, since analog days are only found for the limited region that is positively correlated with the grid cell being downscaled to, LOCA can handle domains that are so large as to include locations where weather systems are independent of each other. In principle one could downscale the entire globe with LOCA given the necessary observationally based training data.

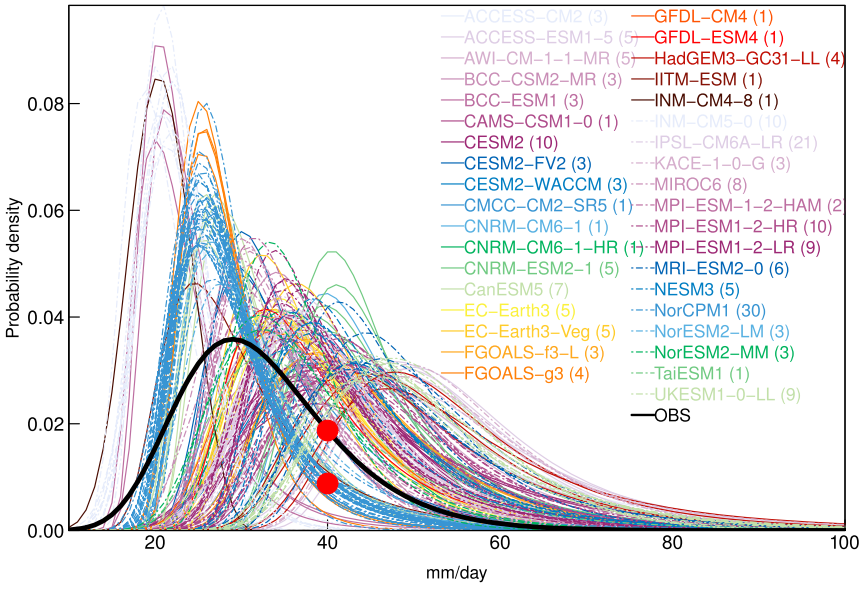

FIG. 1. Probability density functions (PDFs) of the best-fit generalized extreme value (GEV) distributions of extreme daily precipitation (mm day^{-1}) at Seattle, WA, from a variety of CMIP6 models, indicated by the different colors. Observations are shown by the thick black line. The model legend shows, in parentheses, the number of ensemble members plotted for each model. The red dots show, as an example, how NorCPM1 (dot–dashed light blue lines) has about half the probability of experiencing a 40 mm day^{-1} precipitation event as is observed.

b. Precipitation training data

LOCA version 1 used the Livneh et al. (2015, hereafter L15) extended North American data as the training data for daily Tmin, Tmax, and precipitation. However, subsequent comparison of L15 precipitation to station data showed an overly weak representation of daily precipitation extremes (e.g., Pierce et al. 2021; Risser et al. 2021). Pierce et al. (2021) used the Variable Infiltration Capacity (VIC; Liang et al. 1994) land surface/hydrological model to show that the diminution in extreme daily precipitation found in L15 resulted in about a 15% lower overall runoff averaged across the CONUS on an annual time scale, an interesting example of effects across time scales that arises from the nonlinear response of surface runoff generation to precipitation (e.g., Liang et al. 1994). In brief, L15 split precipitation gauge readings from non-midnight-observing stations across two days. This process involves binning and temporal averaging of observations, which diminishes extremes, increases the number of wet days, alters the distribution of wet-spell length, and decreases the average precipitation on wet days (Pierce et al. 2021). To correct this shortcoming, in LOCA2 we use the 6-km gridded daily precipitation dataset of Pierce et al. (2021) as the training data, an updated version of L15 that more realistically preserves daily precipitation extremes and extends the period of record through 2018.

c. Ensemble bias correction

All ESMs have systematic errors, known as biases, which can be problematic for stakeholders who need unbiased data to drive impact models. An example of biases in extreme daily precipitation in the CMIP6 GCMs is given in Fig. 1, which shows the probability density functions (PDFs) of best-fit generalized extreme value (GEV) distribution fits for extreme daily precipitation at Seattle, Washington, over the historical period (1950–2014). Different ensemble members of the same model (multiple lines with the same color) produce different estimates of the GEV due to sampling variability and natural internal climate variability, such as fluctuations associated with El Niño or the Pacific decadal oscillation. Some ESMs underestimate extremes by 50%. For example, in Fig. 1 the upper red dot shows the probability of having a 40 mm day^{-1} event in Seattle from the observations and the lower red dot shows the probability in NorCPM1, which is about half the observed value. Other models overestimate the chance of extreme events by 100% or more. In general, different ensemble members from the same model cluster more closely together than do different GCMs. The goal of our bias correction approach is to reduce systematic biases such as illustrated here while retaining the original GCM's daily weather variability and projected climate change signal.

Many bias correction schemes are based on quantile mapping (QM), which replaces a model value at some percentile with the observed value at that percentile. In its simplest form, QM results in the bias corrected mean becoming equal to the observed mean and model maximum value equal to the observed maximum value. This may be reasonable for a single model run, but it is problematic when applied to multiple ensembles from a single model. One issue is that bias correcting all ensemble members so that they have the same mean

TABLE 1. The 16 stations used here, their station IDs, three letter codes, latitudes, longitudes, elevations (m), U.S. states, and names.

GHCN ID	Code	Lat (°)	Lon (°)	Elev (m)	State	Name
USW00024233	SEA	47.444 4	−122.313 9	112.8	WA	SEATTLE TACOMA INTL AP
USC00240802	BIL	45.771 7	−108.481 1	944.0	MT	BILLINGS WTP
USW00014914	FAR	46.925 3	−96.811 1	274.3	ND	FARGO HECTOR INTL AP
USW00014739	BOS	42.360 6	−71.010 6	3.7	MA	BOSTON LOGAN INTL AP
USW00024131	BOI	43.566 7	−116.240 6	857.7	ID	BOISE AIR TERMINAL
USW00023062	DEN	39.763 3	−104.869 4	1611.2	CO	DENVER-STAPLETON
USC00111577	CHI	41.737 2	−87.777 5	189.0	IL	CHICAGO MIDWAY AP 3SW
USW00013743	DCA	38.848 3	−77.034 2	3.0	VA	WASHINGTON REAGAN AP
USW00023234	SFO	37.658 1	−122.437 8	2.4	CA	SAN FRANCISCO WSO AP
USW00023169	LAS	36.071 9	−115.163 3	649.5	NV	LAS VEGAS MCCARRAN AP
USW00003947	MCI	39.297 2	−94.730 6	306.3	MO	KANSAS CITY INTL AP
USW00013874	ATL	33.630 0	−84.441 7	307.8	GA	ATLANTA HARTSFIELD INTL AP
USW00023188	SAN	32.733 6	−117.183 1	4.6	CA	SAN DIEGO WSO AP
USW00023044	ELP	31.811 1	−106.375 8	1194.2	TX	EL PASO INTL AP
USW00003927	DFW	32.897 8	−97.018 9	170.7	TX	DALLAS FT WORTH AP
USW00012839	MIA	25.790 6	−80.316 4	8.8	FL	MIAMI INTL AP

reduces the total ensemble's natural variability (Chen et al. 2019; Vaittinada Ayar et al. 2021), which can cause appreciable variation even on multidecadal time scales (Deser et al. 2012).

Of more immediate concern is that having all the ensemble members exhibit the same maximum value, which is equal to the observed maximum value, underestimates expected extremes in a dataset where models have multiple ensemble members. For example, several CMIP6 models have 10 or more ensemble members. All else being equal, a model with 10 ensemble members would be likely to have larger extreme values present than a model that only supplies one ensemble member solely due to the larger sample size. Bias correcting all ensemble members so they have identical maxima is not physically realistic.

These general concepts are illustrated at selected cities (Table 1; Fig. 2), chosen to sample a variety of regional climates across the CONUS, in Fig. 3. Although it is an imperfect analogy for reasons discussed below, shown are extreme value plots for annual maximum daily precipitation over the historical period from L15 and three multimodel statistically downscaled products: LOCA version 1, multivariate adaptive constructed analogs (MACA; Abatzoglou and Brown 2012), and bias correction with spatial disaggregation (BCSD; Wood et al. 2002; U.S. Bureau of Reclamation 2013). Both MACA and BCSD use QM to bias correct data over the historical period. MACA is similar to LOCA in some respects, such as downscaling based on constructed analogs, the ability to downscale variables jointly by finding analog days that match more than one variable simultaneously, and taking explicit consideration of how to bias correct projected future model output. They differ in how the analog days are constructed, with MACA constructing analog days over the entire domain while LOCA uses a localized approach that chooses analogs only over the regional synoptic scale that influences a point being downscaled to. The bias correction of future projections also differs in important ways, with MACA using an epoch adjustment that removes the climate trend before downscaling,

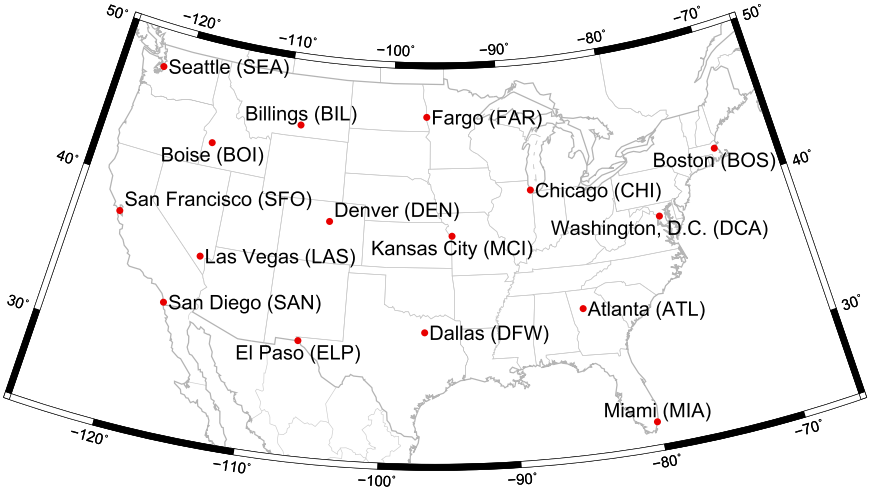

FIG. 2. Location of the 16 stations used in this work, along with the 3-letter city codes used here. See also Table 1.

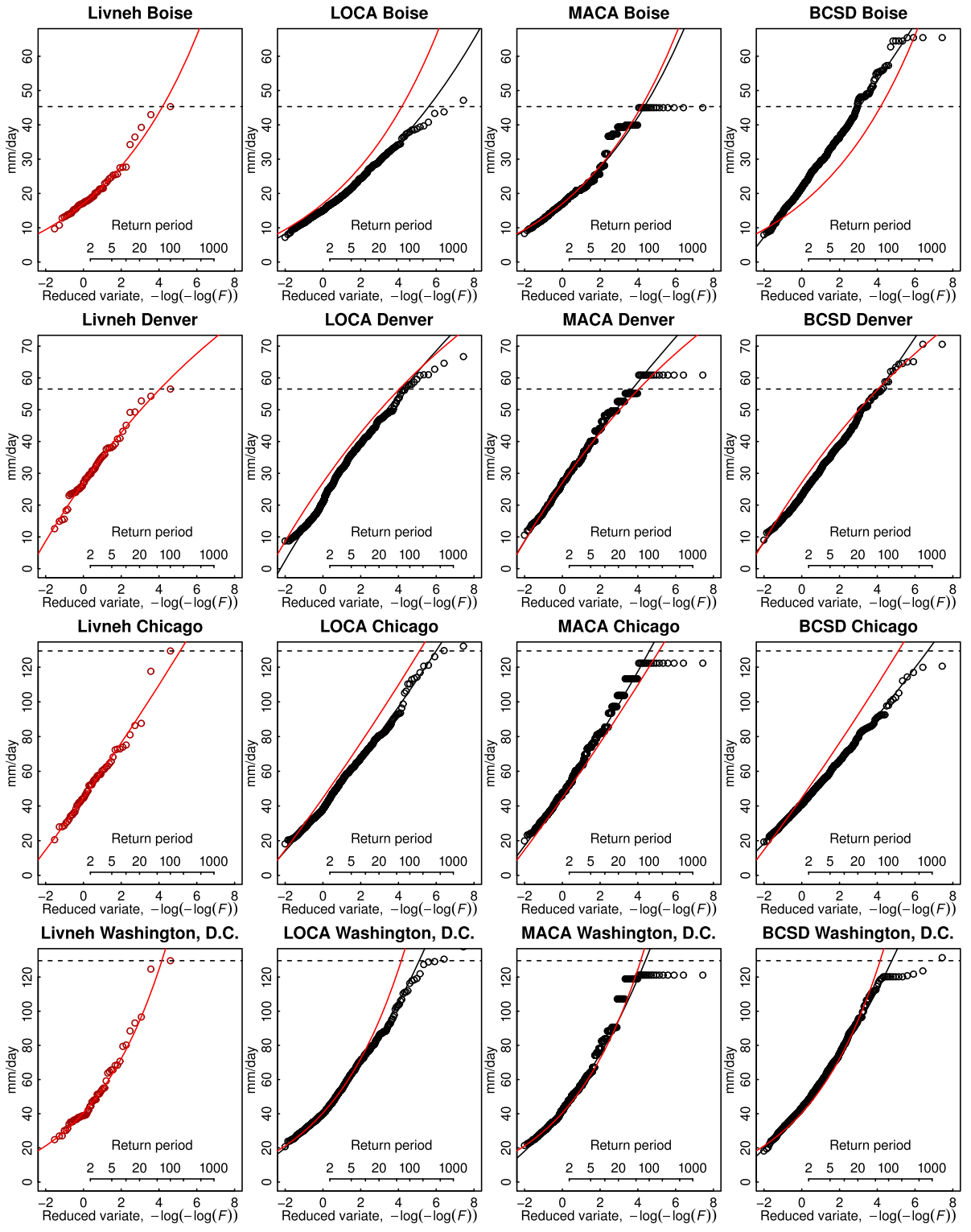

FIG. 3. (first column) An extreme value (reduced variate) plot showing annual block maximum of daily precipitation (red circles; mm day^{-1}) from the L15 dataset. The return period (years) is shown in the inset axis. The red curve shows the best GEV fit, and the horizontal dashed line is the all-time maximum observed value. Each row shows a different city, as indicated in the title. Columns show similar results for all the models in three statistically downscaled multimodel datasets: (second column) LOCA version 1, (third column) MACA (Abatzoglou and Brown 2012), and (fourth column) BCSD (REF). For the statistically downscaled products, the black curve shows the best GEV fit for the model data while the red curve is the best GEV fit to the Livneh observations, repeated from the first column.

while LOCA uses the PresRat technique (Pierce et al. 2015) that preserves the model-predicted future change by quantile. BCSD is less similar to MACA and LOCA as it is not a constructed analog technique and operates on monthly data, choosing an observed historical month as the basis for the downscaling. However, it uses QM bias correction, as do MACA and LOCA, albeit implemented in monthly windows rather than the iterative set of windows LOCA uses (Pierce et al. 2015).

Caution is needed when interpreting Fig. 3; it is not a direct illustration of the bias correction issues described above since the datasets only downscaled one ensemble member from each model and our discussion concerns models with multiple ensemble members. Instead, results are shown across all models in the dataset, comprising 32, 18, and 25 GCMs for LOCA v1, MACA, and BCSD, respectively. Nonetheless some key concepts noted above can be seen. For example, in MACA all the models have the same all-time maximum value, since QM replaces the model's maximum value with the observed maximum value. (The MACA maximum is not necessarily equal to the L15 maximum shown in the figure because MACA used a different observed training dataset than L15.) LOCA version 1 does not have the same kind of hard ceiling across models because the sampling technique used there attempted to reduce this problem by sampling later parts of the historical period together with the early part of the projection period (see Pierce et al. 2015 for details), but the highest values fall systematically below the best-fit GEV (black line) so the attempt to preserve extremes this way was not completely successful. BCSD, unlike LOCA and MACA, is a monthly technique that chooses an analog month to produce daily values, but even so Fig. 3 shows that there is an influence of the bias correction on the daily values that gives a tendency toward a similar (although not identical) maximum value across different models.

Bias correcting all ensemble members together rather than bias correcting each ensemble member individually better preserves variability and extremes across the ensemble members (Chen et al. 2019) and avoids truncating all extremes to the maximum observed value. Models with multiple ensemble members have rarer extremes than are available from the limited observations, so for bias correction purposes we fit a GEV distribution to the observations and use the GEV estimate when bias correcting model points that fall off the end of the observed distribution (as was done in, e.g., Wood et al. 2004). We fit the GEV using block maxima and the method of L-moments (Hosking 1990), chosen because it is computationally fast and performs well with limited data (Hosking 1990).

The LOCA version 2 implementation of ensemble QM bias correction over the historical period is illustrated via a flowchart in Fig. 4. At each spatial location, we start with the time series of the observations as well as the time series of all model ensemble members pooled together over the historical period. The cumulative distribution function of the observations is fitted to a monotonic cubic spline (Fritsch and Carlson 1980), which is used for QM bias correcting nonextreme values. Additionally, the observed block maxima are fit to a GEV using L-moments, which is used for QM bias correcting extremes. We estimate the uncertainty in the three fitted GEV parameters (location, scale, and shape) using the parametric Monte Carlo approach of Kyselý (2008) with 100 simulations. This yields a distribution of GEV parameters estimated from observed extremes. For each model data point that falls outside of the observed distribution, we randomly pick values of the location, scale, and shape parameters consistent with the Monte Carlo parameter distributions and use the resultant GEV to provide an estimate of observed values that are more extreme than can be found in the available observations.

Bias correction over the future period is unchanged from LOCA version 1, which uses the PresRat scheme (Pierce et al. 2015). Briefly, the model-predicted future change ratio (i.e., future value/historical value) is calculated in quantiles, then applied to the bias corrected historical value at each quantile. This preserves the ratio of model calculated future changes by quantile, hence the name PresRat. The windowing approach, frequency-dependent bias correction, and correction of wet day fraction are unchanged from version 1 and described in Pierce et al. (2015).

Example results from the LOCA2 downscaled data over the historical period are illustrated in Fig. 5, which shows extreme value plots of daily precipitation at the 16 airport locations across the continental United States shown in Fig. 2 and Table 1. At each location, results from one randomly selected (without replacement) CMIP6 model with multiple ensemble members are shown. The goal of the ensemble bias correction is to avoid a fixed maximum value of precipitation across all the ensemble members. Figure 5 shows that this goal is largely achieved, with the LOCA2 extreme values for most locations increasing as one would expect given the best-fit GEV (red line) and the uncertainty of the fit (dashed green lines show the 95% confidence interval). ACCESS-ESM1-5 at Miami (MIA), which has 5 ensemble members, shows the most divergence of extreme values from the expected range, so agreement with the expectation from the GEV is not complete. Other models show similar shortfalls in the most extreme values at MIA; we speculate this may be due to deficiencies in the simulation of strong hurricanes in the CMIP6 models. On the other hand, both CESM2-LENS and IPSL-CM6A-LR have 10 ensemble members and agree well with the best-fit GEVs at Fargo, North Dakota (FAR), and Boise, Idaho (BOI), respectively. It is difficult to separate out the effects of sampling and natural variability on the results at these extreme values, which fall at about the 1 day in several centuries level and are subject to considerable uncertainty. One advantage of the LOCA version 2 approach of downscaling up to 10 ensemble members per model is that this provides more samples of rare events for models that provided multiple ensemble members, which reduces sampling uncertainty compared to only having one ensemble member available. In any event, Fig. 5 shows that the ensemble bias correction avoids having all of a model's ensemble members exhibit the same maximum value, and so does not give a physically unrealistic projection.

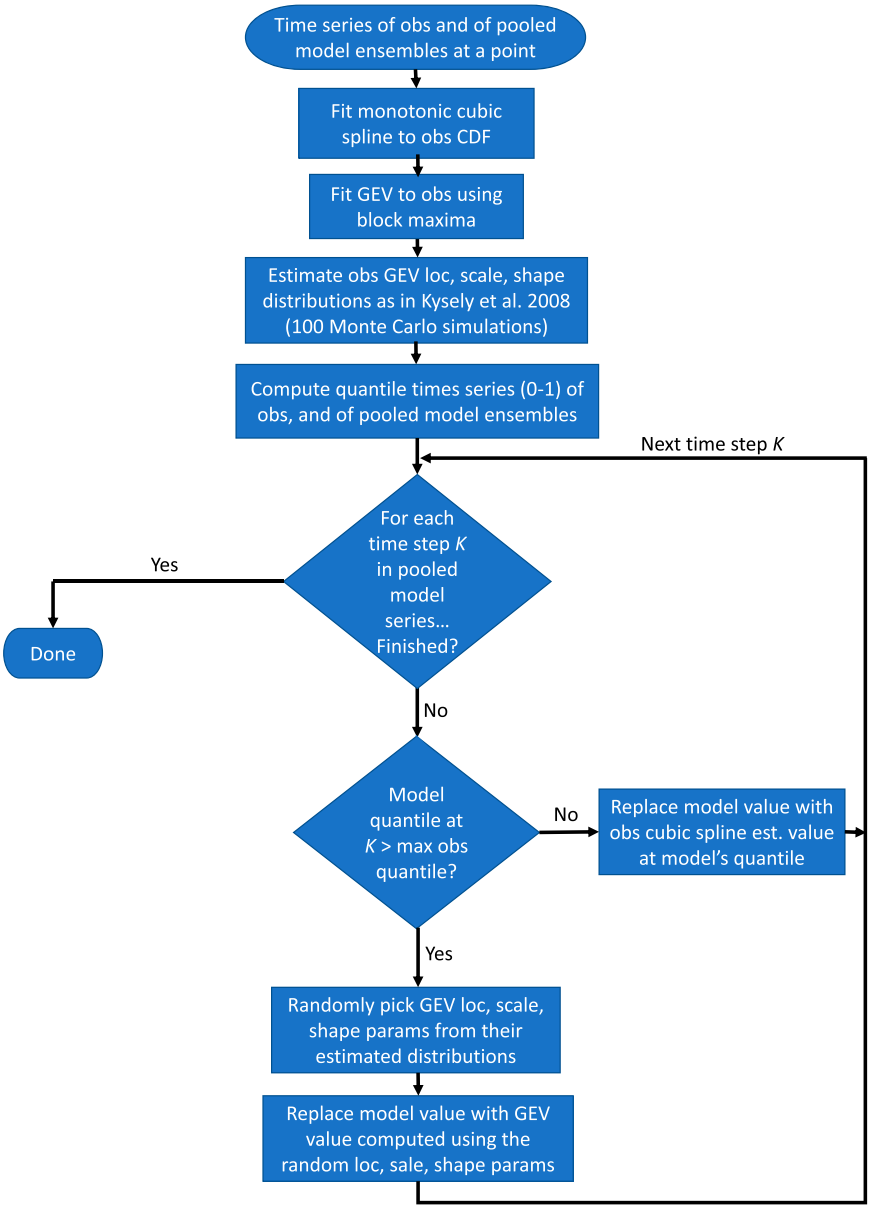

FIG. 4. Flowchart of the LOCA procedure for constructing the bias corrected and GEV extrapolated extreme model values illustrated in Fig. 5.

d. *Observational data errors and extreme precipitation limits*

Station observations occasionally have errors that lead to unrealistic GEV fits in the ensemble bias correction approach, producing unrealistic extremes in the downscaled result. As an example, on 3 May 1980 a total of 656 mm of precipitation is reported in Sinaloa, Mexico (26.78°N, 109.41°W). We deem it an error since it is more than an order of magnitude larger than the second largest May precipitation value and the surrounding locations were dry on that day. Although we addressed this instance by setting the value to zero, a more general guard against such observational errors is needed, especially where erroneous values in the observational record are less readily apparent.

To address this issue, we implement a spatially varying limit on extreme daily precipitation values (Fig. 6), while being sensitive to ensuring that actual extreme precipitation values are not inadvertently discarded. The top panel shows the maximum value of precipitation from the training data for reference. Models can have up to 10 ensemble members run over the period 1950–2100, yielding 1500 years of data, and up to three different emissions scenarios, potentially giving 4500 years of data for a model. This motivated using the estimated 5000-yr return value, derived from the GEV parameter fitting described above, as the basis of the mask (middle panel). Sampling variability and uncertainty in GEV parameter fitting produces a spatially noisy mask, which needs to be

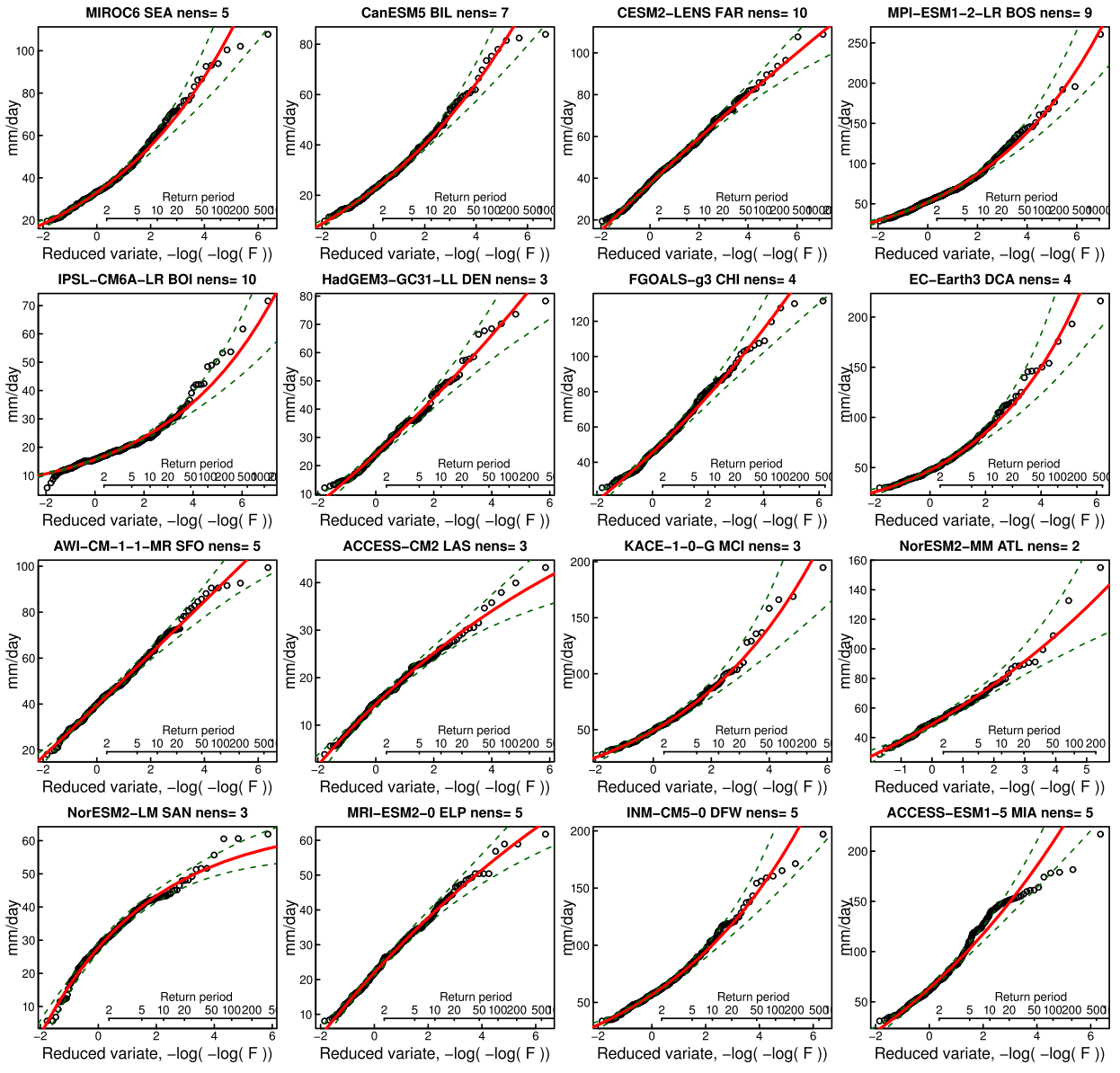

FIG. 5. Extreme value plots of maximum daily precipitation over the historical period (1950–2014) from LOCA version 2 at the 16 stations shown in Fig. 2. Panels are arranged roughly geographically. The inset axis shows the return period (years). For each station, results from one randomly selected model (without repeats) that has multiple ensemble members are shown; the number of ensemble members is shown in the title. The red line is the best-fit GEV, and the dashed green lines are the limits of the 95% confidence interval of the fitted GEV.

smoothed. As a simple approach we applied a Gaussian spatial smoother with a 2σ radius length scale of 167 km (~28 grid cells) east of 113°W and 50 km (~8 grid cells) west of 115°W. The smaller radius in the western United States avoids smoothing away the high return values in the Sierra Nevada and generally acknowledges the impact of highly variable topography on extreme precipitation in the western part of the country. Between 113° and 115°W, the smoothing radius was linearly switched between the two endpoints. Smoothing decreases the peak return values, so to restore them we multiplied the final field by 2.5, a value selected to allow only 1% of the grid cells to have a maximum mask value less than the 5000-yr return value. The final result (bottom panel of Fig. 6) shows the limits are largest in hurricane prone areas along the Atlantic coast and Gulf of Mexico, as well in the Sierra Nevada and Baja Peninsula. This approach produces a practical mask for erroneous observations that could otherwise bias extreme precipitation solutions in the downscaled data.

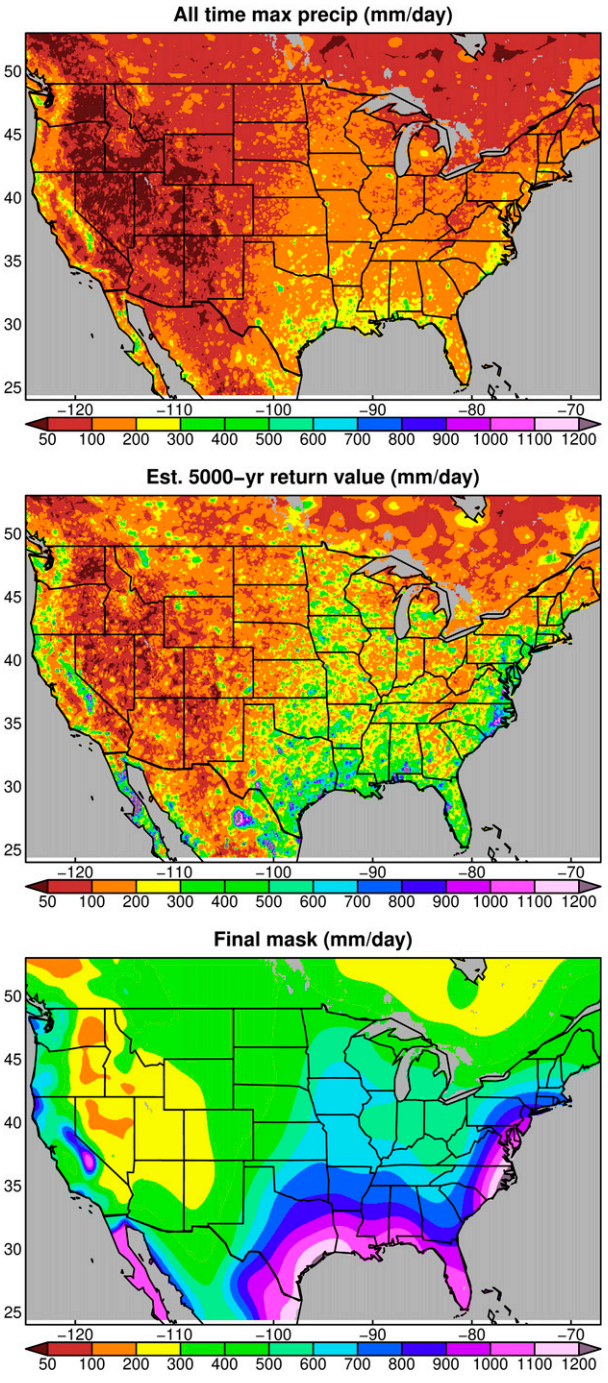

FIG. 6. (top) All-time maximum daily precipitation (mm) from the Pierce et al. (2021) dataset, 1950–2014. (middle) Estimated 5000-yr return value of maximum daily precipitation. (bottom) Final maximum precipitation mask. See text for construction details.

e. *CMIP6 models*

Data from the latest version of the international ESM archive, CMIP6, are used as the source data to be statistically downscaled with LOCA version 2. This includes far more data than produced from the CMIP5 models in the LOCA version 1 effort. LOCA version 2 contains more emissions scenarios, which have been reconfigured as shared socioeconomic pathways (SSPs; O'Neill et al. 2014, 2016; Riahi et al. 2017); up to 10 ensemble members for SSPs 245, 370, and 585 were downscaled (roughly, medium, medium-high, and high emissions scenarios). LOCA version 1 included data from one ensemble member for 32 CMIP5 models, while LOCA version 2 includes data from 27 CMIP6 models. This reduction in the number of models arises because the GCM data necessary for our downscaling objectives, which is future projections tied to specific emissions scenarios (SSP245, 370, or 585) are not required experiments for inclusion in CMIP6. The full set of GCMs, SSPs, and ensemble members included in LOCA version 2 is shown in Table 2. It includes all GCMs that were available to us with the required data as of mid-2021. Counting each combination of model, experiment (SSP), and ensemble member as a separate run, there are a total of 329 runs, constituting 26 026 years of model data.

3. Results

Multimodel ensemble average (MMEA) projected changes in the 5-yr return value of daily precipitation are shown in Fig. 7. Here and below the MMEA is calculated by first averaging the projected changes across all the ensemble members with the target SSP for each model, then averaging across the models. This gives each model the same weight in the final result. Only locations where at least two-thirds of the models agree on the sign of the projected change are shown in color. Recent work has suggested that pooling data across all model ensembles may give more robust estimates of future changes in precipitation intensity, duration, and frequency than the more traditional MMEA approach used here (Srivastava et al. 2021) but this approach was not implemented in the current work.

Most of the changes in precipitation extremes are positive, though results vary somewhat across seasons and location. Generally, winter increases of 10%–30% in 5-yr daily extremes are seen across most of the CONUS and Canada. The exception is in Mexico (MX), Arizona (AZ), New Mexico (NM), and parts of West Texas, which show little to no increase in the 5-yr return value. Decreases are seen in central Mexico. The latter "dry" region intensifies and expands in the spring, with parts of Mexico showing a greater than 40% decline in 5-yr return value. Projected changes are the smallest in summer and show least agreement across models in sign of the change, with weak indications of lower return values over the lower Colorado River basin and parts of Northern California. New England (the states of Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, and Connecticut) is the exception to the summer result, showing model-consistent increases in all seasons. The annual change shows no consistent sign (positive or negative) across the models in AZ, NM, and MX, and 10%–20% increases across the rest of the domain.

Changes in the 50-yr return value in daily precipitation, shown in Fig. 8, generally amplify the increases found for the 5-yr return period. In comparison to the 5-yr return value (Fig. 7), there is less model agreement on the winter decrease in MX and AZ, although some values are still weakly

TABLE 2. CMIP6 model, experiment (emissions scenario), and ensemble members that were downscaled with LOCA version 2.

1	ACCESS-CM2	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1
2	ACCESS-CM2	ssp245	r1i1p1f1, r2i1p1f1, r3i1p1f1
3	ACCESS-CM2	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1
4	ACCESS-CM2	ssp585	r1i1p1f1, r2i1p1f1, r3i1p1f1
5	ACCESS-ESM1-5	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
6	ACCESS-ESM1-5	ssp245	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
7	ACCESS-ESM1-5	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
8	ACCESS-ESM1-5	ssp585	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
9	AWI-CM-1-1-MR	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
10	AWI-CM-1-1-MR	ssp245	r1i1p1f1
11	AWI-CM-1-1-MR	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
12	AWI-CM-1-1-MR	ssp585	r1i1p1f1
13	BCC-CSM2-MR	historical	r1i1p1f1
14	BCC-CSM2-MR	ssp245	r1i1p1f1
15	BCC-CSM2-MR	ssp370	r1i1p1f1
16	BCC-CSM2-MR	ssp585	r1i1p1f1
17	CESM2-LENS	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
18	CESM2-LENS	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
19	CNRM-CM6-1	historical	r1i1p1f2
20	CNRM-CM6-1	ssp245	r1i1p1f2
21	CNRM-CM6-1	ssp370	r1i1p1f2
22	CNRM-CM6-1	ssp585	r1i1p1f2
23	CNRM-CM6-1-HR	historical	r1i1p1f2
24	CNRM-CM6-1-HR	ssp585	r1i1p1f2
25	CNRM-ESM2-1	historical	r1i1p1f2
26	CNRM-ESM2-1	ssp245	r1i1p1f2
27	CNRM-ESM2-1	ssp370	r1i1p1f2
28	CNRM-ESM2-1	ssp585	r1i1p1f2
29	CanESM5	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1
30	CanESM5	ssp245	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1
31	CanESM5	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1
32	CanESM5	ssp585	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1
33	EC-Earth3	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1
34	EC-Earth3	ssp245	r1i1p1f1, r2i1p1f1, r4i1p1f1
35	EC-Earth3	ssp370	r1i1p1f1, r4i1p1f1
36	EC-Earth3	ssp585	r1i1p1f1, r3i1p1f1, r4i1p1f1
37	EC-Earth3-Veg	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
38	EC-Earth3-Veg	ssp245	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
39	EC-Earth3-Veg	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1
40	EC-Earth3-Veg	ssp585	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1
41	FGOALS-g3	historical	r1i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
42	FGOALS-g3	ssp245	r1i1p1f1, r3i1p1f1, r4i1p1f1
43	FGOALS-g3	ssp370	r1i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
44	FGOALS-g3	ssp585	r1i1p1f1, r3i1p1f1, r4i1p1f1
45	GFDL-CM4	historical	r1i1p1f1
46	GFDL-CM4	ssp245	r1i1p1f1
47	GFDL-CM4	ssp585	r1i1p1f1
48	GFDL-ESM4	historical	r1i1p1f1
49	GFDL-ESM4	ssp245	r1i1p1f1
50	GFDL-ESM4	ssp370	r1i1p1f1
51	GFDL-ESM4	ssp585	r1i1p1f1
52	HadGEM3-GC31-LL	historical	r1i1p1f3, r2i1p1f3, r3i1p1f3
53	HadGEM3-GC31-LL	ssp245	r1i1p1f3
54	HadGEM3-GC31-LL	ssp585	r1i1p1f3, r2i1p1f3, r3i1p1f3
55	HadGEM3-GC31-MM	historical	r1i1p1f3, r2i1p1f3
56	HadGEM3-GC31-MM	ssp585	r1i1p1f3, r2i1p1f3
57	INM-CM4-8	historical	r1i1p1f1
58	INM-CM4-8	ssp245	r1i1p1f1
59	INM-CM4-8	ssp370	r1i1p1f1

TABLE 2. (*Continued*)

60	INM-CM4-8	ssp585	r1i1p1f1
61	INM-CM5-0	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
62	INM-CM5-0	ssp245	r1i1p1f1
63	INM-CM5-0	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
64	INM-CM5-0	ssp585	r1i1p1f1
65	IPSL-CM6A-LR	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
66	IPSL-CM6A-LR	ssp245	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
67	IPSL-CM6A-LR	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
68	IPSL-CM6A-LR	ssp585	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1
69	KACE-1-0-G	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1
70	KACE-1-0-G	ssp245	r1i1p1f1, r2i1p1f1, r3i1p1f1
71	KACE-1-0-G	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1
72	KACE-1-0-G	ssp585	r1i1p1f1, r2i1p1f1, r3i1p1f1
73	MIROC6	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
74	MIROC6	ssp245	r1i1p1f1, r2i1p1f1, r3i1p1f1
75	MIROC6	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1
76	MIROC6	ssp585	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
77	MPI-ESM1-2-HR	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
78	MPI-ESM1-2-HR	ssp245	r1i1p1f1, r2i1p1f1
79	MPI-ESM1-2-HR	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r9i1p1f1, r10i1p1f1
80	MPI-ESM1-2-HR	ssp585	r1i1p1f1, r2i1p1f1
81	MPI-ESM1-2-LR	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r10i1p1f1
82	MPI-ESM1-2-LR	ssp245	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r10i1p1f1
83	MPI-ESM1-2-LR	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r7i1p1f1, r8i1p1f1, r10i1p1f1
84	MPI-ESM1-2-LR	ssp585	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1, r7i1p1f1, r8i1p1f1, r10i1p1f1
85	MRI-ESM2-0	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
86	MRI-ESM2-0	ssp245	r1i1p1f1
87	MRI-ESM2-0	ssp370	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
88	MRI-ESM2-0	ssp585	r1i1p1f1
89	NorESM2-LM	historical	r1i1p1f1, r2i1p1f1, r3i1p1f1
90	NorESM2-LM	ssp245	r1i1p1f1, r2i1p1f1, r3i1p1f1
91	NorESM2-LM	ssp370	r1i1p1f1
92	NorESM2-LM	ssp585	r1i1p1f1
93	NorESM2-MM	historical	r1i1p1f1, r2i1p1f1
94	NorESM2-MM	ssp245	r1i1p1f1, r2i1p1f1
95	NorESM2-MM	ssp370	r1i1p1f1
96	NorESM2-MM	ssp585	r1i1p1f1
97	TaiESM1	historical	r1i1p1f1
98	TaiESM1	ssp245	r1i1p1f1
99	TaiESM1	ssp370	r1i1p1f1

negative. Fifty-year daily extremes also increase more over the rest of the CONUS and Canada than 5-yr extremes, although the differences are modest. Parts of the upper Midwest show increases of over 30% in winter and spring, which has serious implications for flooding in the region. Annual changes are more broadly distributed, with increases of 10%–20% widely spread across most of the CONUS and Canada. The model consistency in projected increases in summer precipitation is again seen in New England.

Changes across season, SSP, location, and return value are summarized in Fig. 9, which shows MMEA projected change in return values by the end of the century (2075–2100 with respect to 1950–2014). The top row (solid lines) shows the spatial average across the north-central tier of the domain, including Canada and the CONUS except for NM and AZ. The greatest changes with respect to how extreme the return value is (from 5 to 50 to 500 years, moving from the top-left to the top-right panel) are found in the summer months (June–August). Little change in JJA 5-yr return value is found in any SSP (top-left panel), and what changes are found are not significantly different from zero given the spread across model results (vertical bars to the right of each panel). This is unlike the other three seasons, which show a consistent increase in 5-yr return values as the emissions scenario becomes more

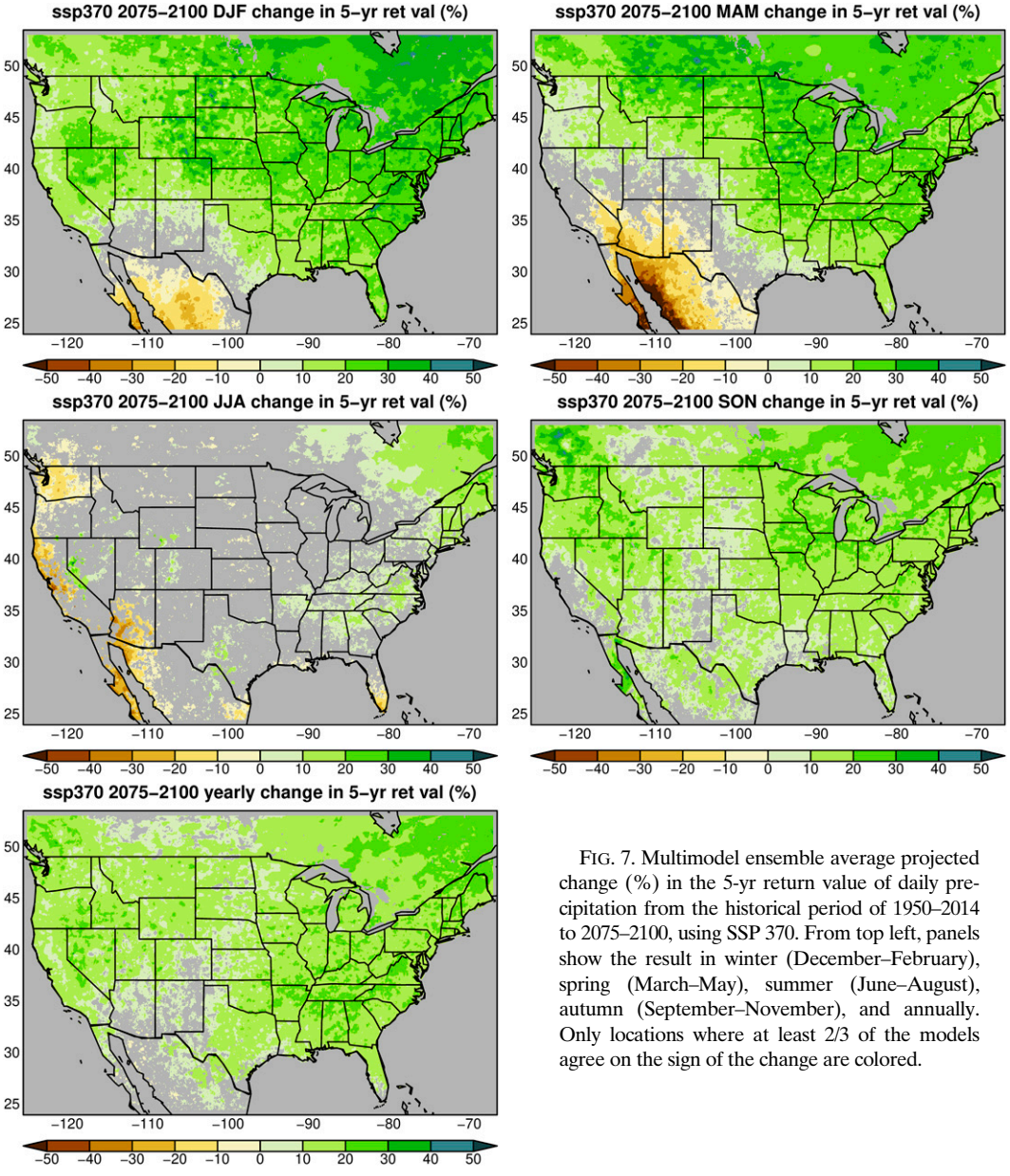

FIG. 7. Multimodel ensemble average projected change (%) in the 5-yr return value of daily precipitation from the historical period of 1950–2014 to 2075–2100, using SSP 370. From top left, panels show the result in winter (December–February), spring (March–May), summer (June–August), autumn (September–November), and annually. Only locations where at least 2/3 of the models agree on the sign of the change are colored.

extreme, with all changes different from zero. Increases are greatest in winter (December–February), which may be of particular importance where flooding is most likely in winter, such as on the west coast.

Moving from the 5- to 50- to 500-yr return values (top-left panel to top-right panel of Fig. 9), summer is again the outlier, showing a much stronger increase at progressively longer return periods than is found in the other seasons. At the longest return period examined here, 500 years, the summer increase in extreme values is roughly 20%, overlapping with the spring increases given the uncertainties. Ignoring summer for the moment, the other three seasons show relatively uniform increases as the emissions scenario increases from SSP 245 to 585 and as the return period increases from 5 to 500 years. Extreme increases are the largest in winter in all cases (i.e., across the different return periods and SSPs), ranging from about 17% for the 5-yr return value under emissions scenario SSP245 to almost 40% for the 500-yr return value in SSP585.

Given that the projected winter increases in the 50-yr return value of daily precipitation are about 25% for the middle emissions scenario considered here, SSP 370, locations where winter flooding is a concern are under considerable threat for increased incidence of damaging floods in coming decades. Even the lowest scenario considered here (SSP 245) shows increases of 18% by the end of this century.

A different picture is found in the southern part of the domain (Mexico, AZ, and NM; middle row of Fig. 9, dashed lines). At the less extreme (5-yr) return values, strong decreases

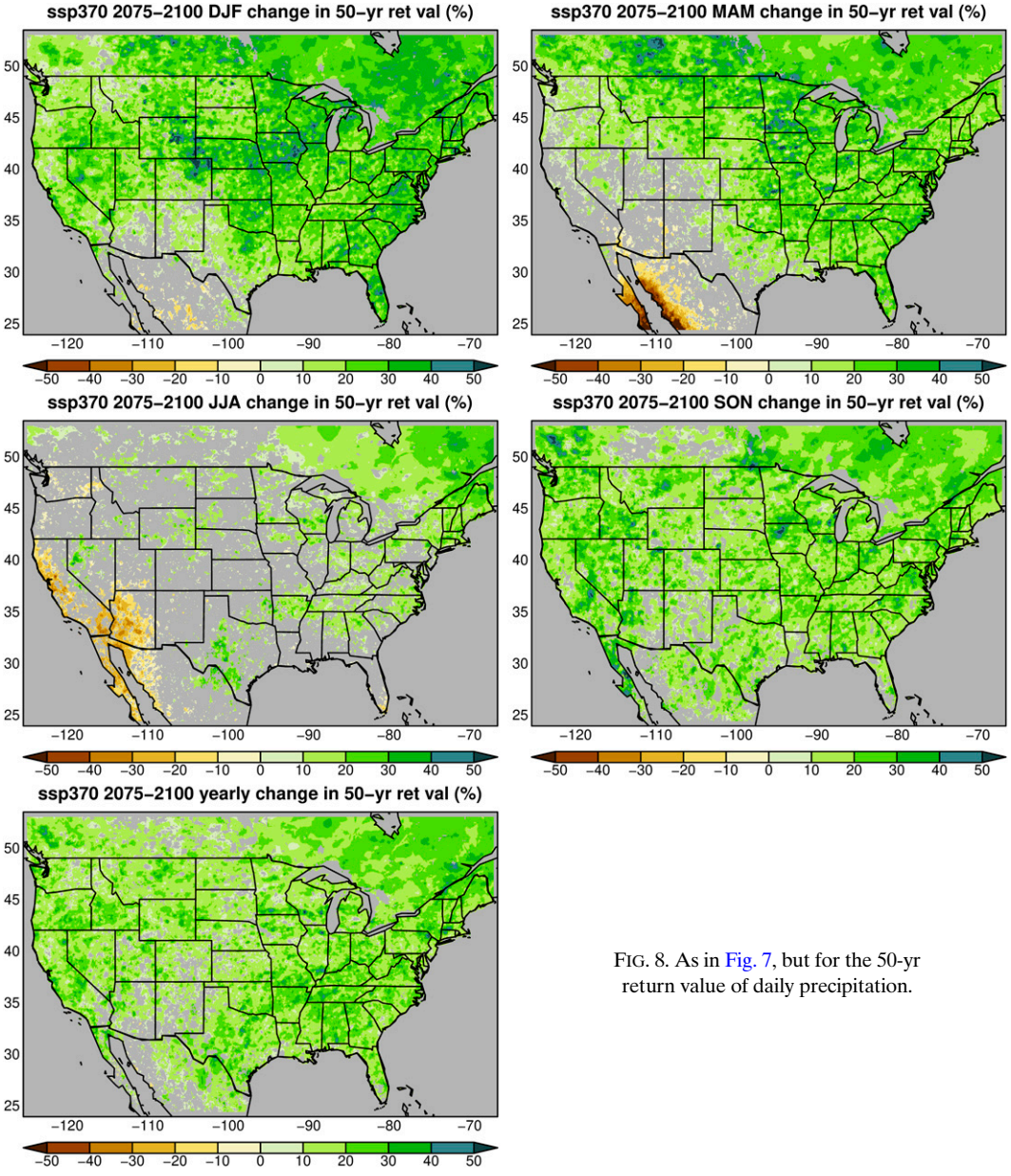

FIG. 8. As in Fig. 7, but for the 50-yr return value of daily precipitation.

are found in spring (March–May), and weak decreases that overlap with zero in DJF and JJA. In this part of the domain the only consistent increases in extreme daily precipitation are found in autumn (September–November), when changes are much more consistent across the domain (top row of Fig. 9 versus bottom row), unlike what is found in the other seasons. At progressively longer return periods (from 5 to 500 years, middle left to middle right) the changes in daily precipitation extremes become larger, until all changes are positive at the 500-yr return value (middle right), although summer changes are not distinct from zero given the uncertainty. Winter shows the largest increases in the northern part of the domain, but autumn shows the largest increase in the southern part with increases ranging from 10% (5-yr return value, SSP 245) to 40% (500-yr return value, SSP 585).

Finally, as noted above, New England (bottom row of Fig. 9) shows a consistent increase in extremes across all seasons including summer, unlike most of the domain. Additionally, at the 500-yr return period New England is projected to experience the largest seasonal increases primarily in summer, unlike the other regions where the increases across seasons are smallest in summer.

a. Change in return period for a given event

Up to now we have focused on the change in value at a given return period. Since the design of water management infrastructure is often informed by return values at specified return periods, these return value changes indicate how future infrastructure designs might be affected by climate change. A complementary approach is to examine how often a historically

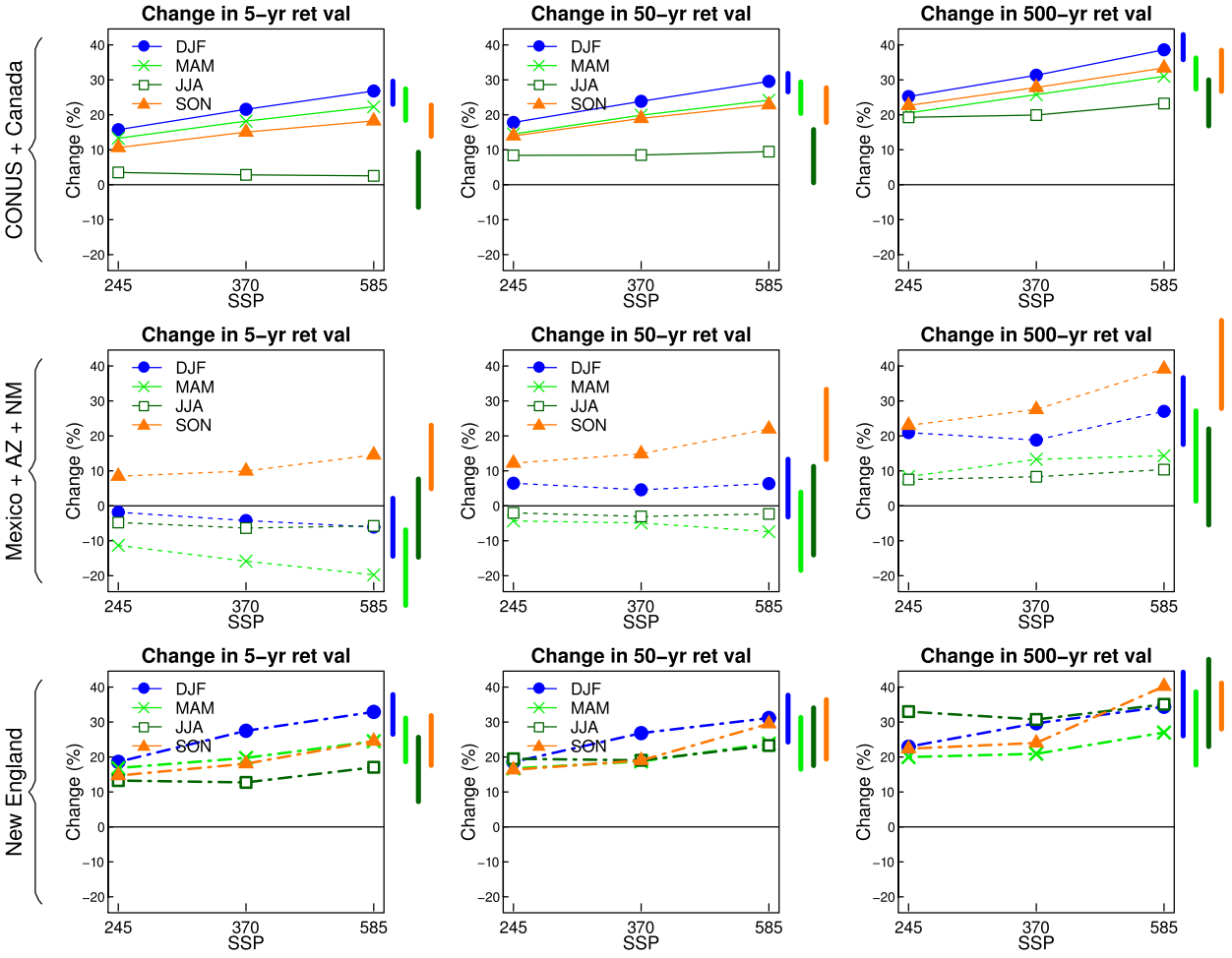

FIG. 9. (top) Solid lines are MMEA projected change (%) in 5-, 50-, and 500-yr return values of daily precipitation as a function of SSP (*x* axis) averaged over Canada and the CONUS exclusive of AZ and NM. Different colored lines are shown for each season, as indicated by the legend. (middle) As in the top row, but dashed lines are averaged over MX, AZ, and NM. (bottom) As in the top row, but dash–dot lines are averaged over New England (Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, and Connecticut). Changes are calculated from the historical period of 1950–2014 to 2075–2100. The vertical bars to the right of the panels show the interquartile ranges across the models for each season as a measure of uncertainty.

extreme event will occur in the future, which has relevance for understanding how existing infrastructure that was designed using historically applicable return values might be stressed by future climate changes.

The estimated future return period (years) for the historically defined 100-yr return value of daily precipitation at each point is shown in Fig. 10 as a function of time horizon (columns) and SSP (rows). Values are calculated as the median across all model ensemble averages that were run for the indicated SSP. Although noisy, a consistent pattern is seen with the most pronounced increases in frequency in the eastern half of the continent, particularly the southeastern United States and eastern Canada, and in the Pacific Northwest. By the end of this century over much of the southeastern United States, a daily precipitation event that historically recurred only about once a century is projected to become much more frequent, with return periods of 30–40 years depending on the emissions scenario. Changes of this magnitude will have serious implications for the reliability or operations of historical water management infrastructure.

Mild increases in return period are seen in the early period (2015–44) over AZ, NM, and parts of MX. These increases mostly disappear by midcentury and have reversed in sign (indicating the more frequent occurrence of historically extreme events) by the end of this century, especially at SSPs 370 and 585.

b. Changes in the distribution of daily precipitation

Our results have shown that the models project a widespread increase in daily precipitation extremes, especially outside of MX, AZ, and NM, with increases larger as the century progresses and at higher emissions levels (SSPs). Are these increases in the extremes simply a manifestation of overall increases in precipitation across a range of magnitudes, or are

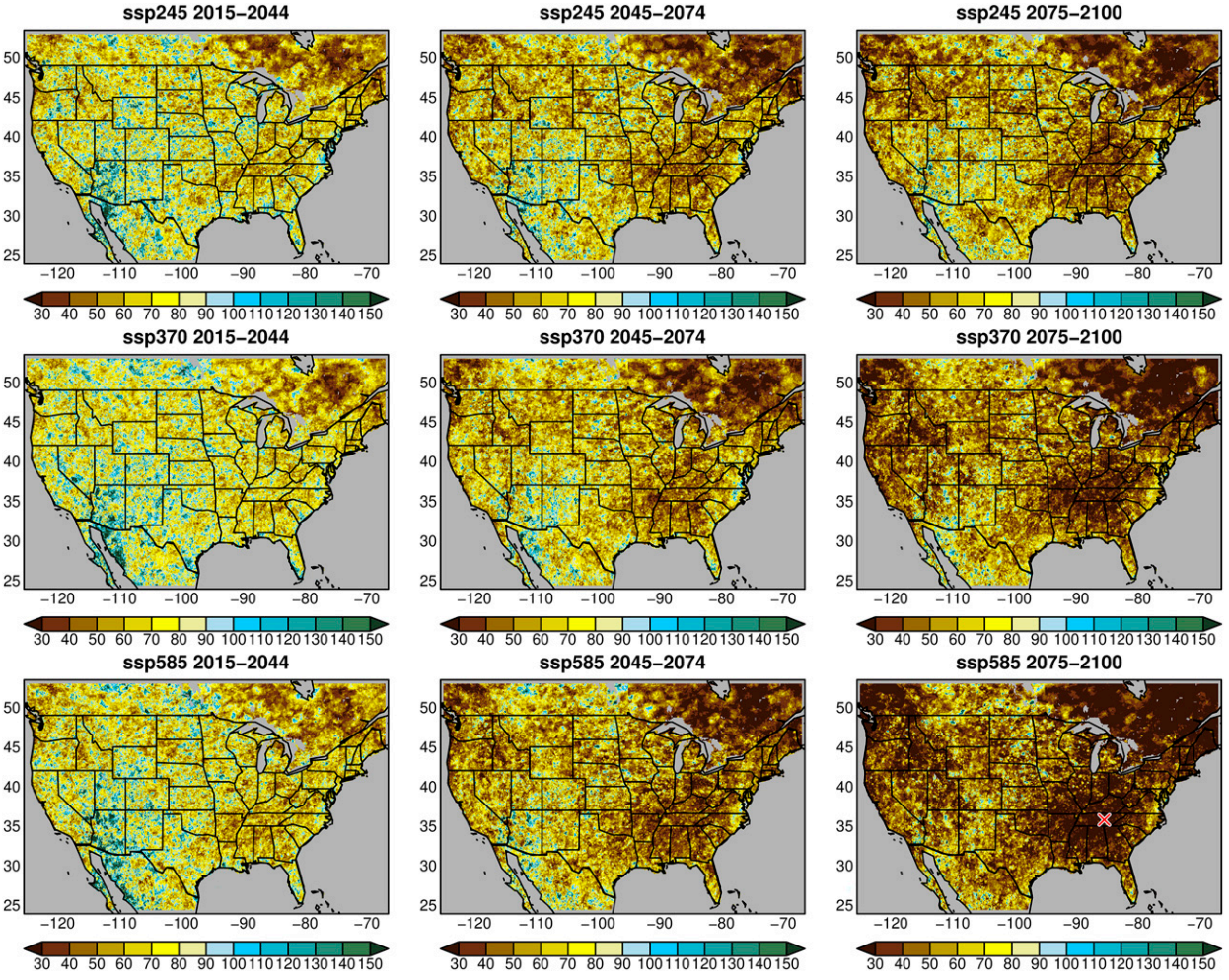

FIG. 10. Future return period (years) for the historical (1950–2014) 100-yr return value of daily precipitation. For example, the red cross over Tennessee in the bottom-right panel shows that a daily precipitation event that was a 100-yr return value in the historical era would be expected to occur about every 30 years by 2075–2100 under SSP 585. Displayed values are the median across all models that have at least one run of the indicated SSP.

they part of a more complicated change in the distribution of daily precipitation values?

Projected changes (2075–2100, SSP 370) in the number of wet days that fall in historically defined percentile bins of wet day precipitation amounts are shown in Fig. 11. There are strong increases in the number of days of the most extreme precipitation (99.9th percentile), typically by a factor of 1.5–3, at all stations except Miami (MIA). Notably, the number of extreme precipitation days is projected to increase even for the southern tier of stations, where annually averaged precipitation and the fraction of wet days are both projected to decrease. Therefore, the increases in extremes are not a simple manifestation of generally increasing precipitation. Looking at the changes in detail, at almost all stations the increase in extreme days is accompanied by a decrease in low-to-mid-percentile wet days (25th–75th percentile). This is seen even at stations with an overall increase in annually averaged precipitation. For example, at Seattle (SEA), annually averaged precipitation is projected to increase by about 5%, but two-thirds of the models agree that the number of wet days in the (historically defined) 25th–75th percentiles decrease, and overall the fraction of wet days decreases by 5% as well. Similar decreases in wet days accompanied by increases in the wettest days have been shown to be important in understanding projected precipitation changes in California (Pierce et al. 2013a; Polade et al. 2017). The present results show the shift to fewer low-precipitation days and more high-precipitation days occurs across the entire CONUS.

c. *Changes in storm characteristics*

Another question of interest is how storms might change in the future. For instance, will intense precipitation storms become larger or last longer? A straightforward way of evaluating this is by selecting days when precipitation at a station is extreme (in the 99.9th percentile of historical wet days), which we call storm days, and examining how extensive precipitation is in the surrounding region, and how long the extreme precipitation lasts before and after the storm day. For example,

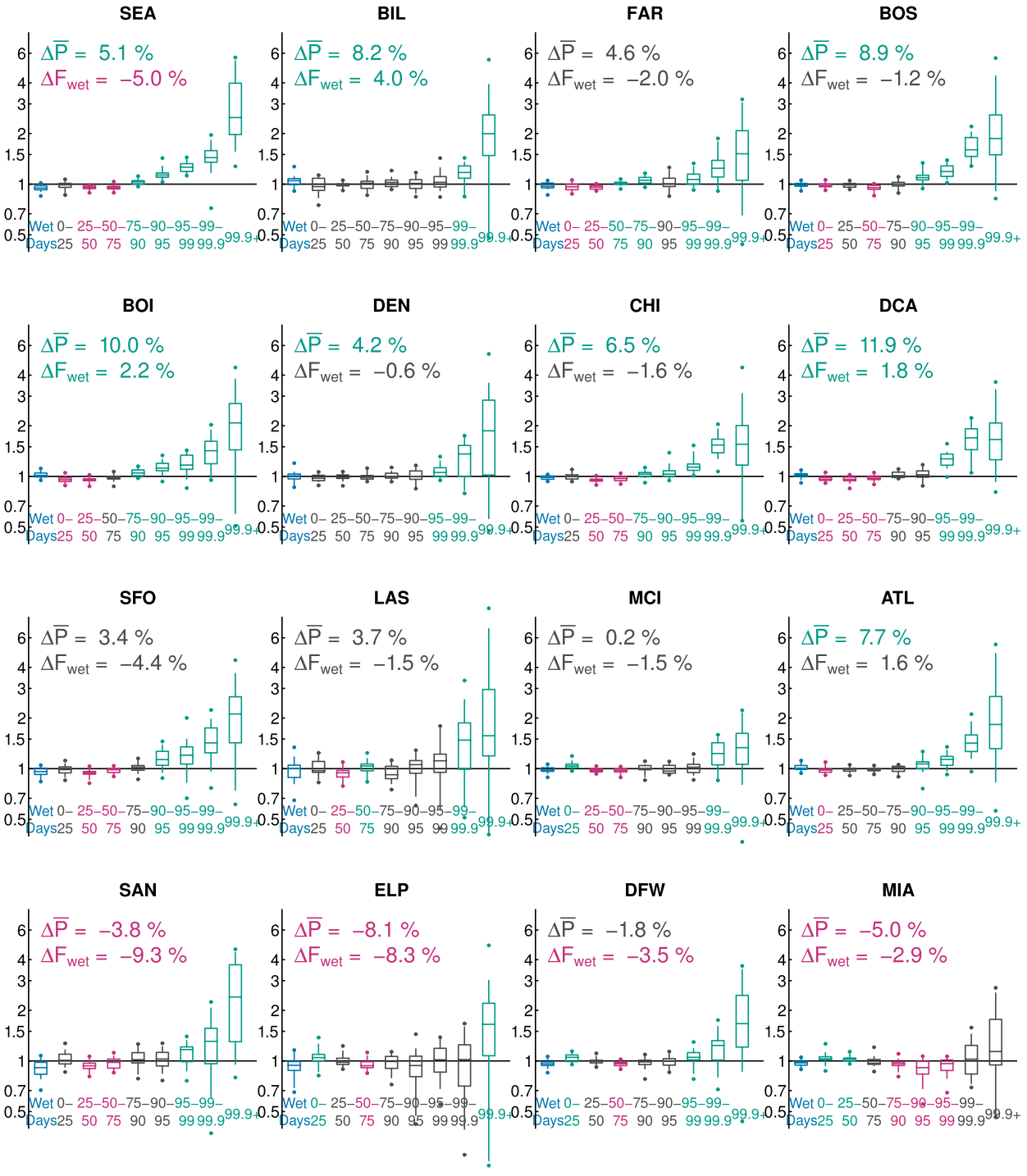

FIG. 11. Model-projected changes (factor) in the number of days with precipitation falling in historically defined wet-day percentile bins at the 16 stations illustrated in Fig. 2, calculated for the end of this century (2075–2100) using SSP 370. Panels are arranged roughly geographically. The box and whiskers illustrate the mean, interquartile range, 5%–95% confidence interval, and extremes (dots) of the distribution of projected changes across models and ensemble members. Teal indicates more than 2/3 of the models agree the change is >1 (wetter), and magenta indicates more than 2/3 of the models agree the change is <1 (drier). Gray indicates lack of agreement across the models on the sign of the change. The blue box at the left shows the change (factor) in number of wet days per year. The notations show the change in annual mean precipitation ($\Delta\overline{P}$) and fraction of wet days (ΔF_{wet}) at the station, with values in teal showing model agreement on increases and magenta showing model agreement on decreases.

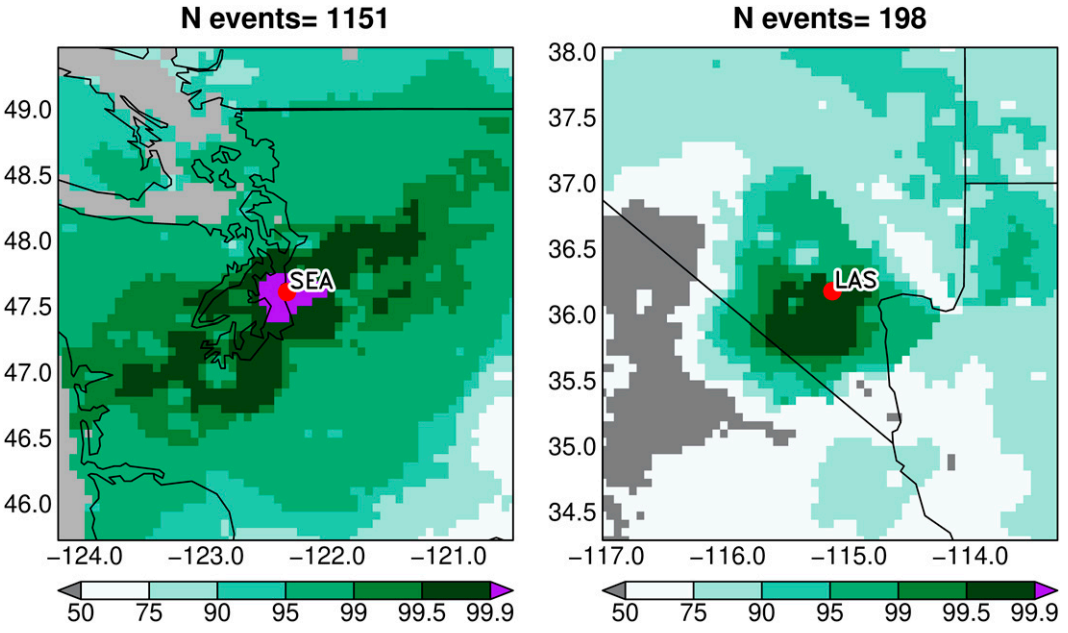

FIG. 12. Multimodel ensemble average precipitation composited on days when precipitation is ≥99.9th percentile of wet days at the indicated station location. Precipitation values are individually normalized at each grid cell to show the local wet day percentile. The panel title shows the total number of extreme wet day events contributing to the composite.

Fig. 12 shows the multimodel ensemble average precipitation on storm days in the region around Seattle (left panel) and Las Vegas, Nevada (LAS; right panel). Since precipitation is strongly affected by topography, values are presented as percentiles normalized to the historical distribution at each grid cell. This allows values to be more sensibly compared across the domain. Since there are considerably more wet days at Seattle than Las Vegas, more extreme values are seen at Seattle. Additionally, Seattle shows a distinct southwest-to-northeast pattern associated with atmospheric rivers, which are the primary source of heavy precipitation events in this region (Ralph et al. 2014). Averaged over the regions shown, values are much larger for Seattle, where storm days are associated with extreme precipitation over almost the entire region. By contrast, storm days at Las Vegas often take the form of convective systems with more regionally confined precipitation patterns.

Constructing these maps for all extreme precipitation events at all 16 stations allows us to see how the return period (1/frequency, in wet days) averaged over the region varies as a function of location, future period, and time offset with respect to the storm day (Fig. 13). Larger values indicate storms with more extreme precipitation. The substantial differences between the storm day historical value at different stations (e.g., 1200 for SEA versus 150 for LAS) are due to the effects already described in Fig. 12, with some locations characterized by extreme days where precipitation is widespread and others where precipitation is more localized. There are also differences in how storm characteristics change in the future. At locations such as Boise; San Francisco, California; and San Diego, California, there are substantial future increases in storm intensity that extend to the day before and after the storm day. By contrast, many of the interior stations show little to no changes in the day before or after the storm day. A few stations, such as Billings, Montana, and Boston, Massachusetts, show less change with future period than found at most other stations.

d. Extreme precipitation compared to LOCA version 1

The better depiction of daily precipitation extremes in LOCA version 2 arises from the improved training data and ensemble bias correction approach. The relative contribution of these two is examined in Fig. 14, which shows the 20-yr return value of daily precipitation from 6 datasets. The top-left panel shows return values calculated directly from daily station data in the Global Historical Climatology Network (GHCNd; Menne et al. 2012). Gridding the 20-yr return values from the stations yields a mean over the CONUS of 97.8 mm day^{-1} (top right). Directly gridding the station return values gives a better representation of station-based data than is found by gridding the daily data and calculating the 20-yr return values from the gridded daily data (e.g., Risser et al. 2021). Essentially, gridding the daily data reduces long-period extremes by combining values from multiple stations when in unobserved locations between stations. Since surrounding stations do not always have their extreme precipitation values fall on the same day, extreme values from one station tend to be diluted by nonextreme values from other nearby stations.

Compared to the directly gridded station return values, L15 (middle left) is weaker along coastal regions of the eastern half of the United States, with a mean of 69.3 mm day^{-1}, 29% less than the directly gridded return values. Pierce et al. (2021)

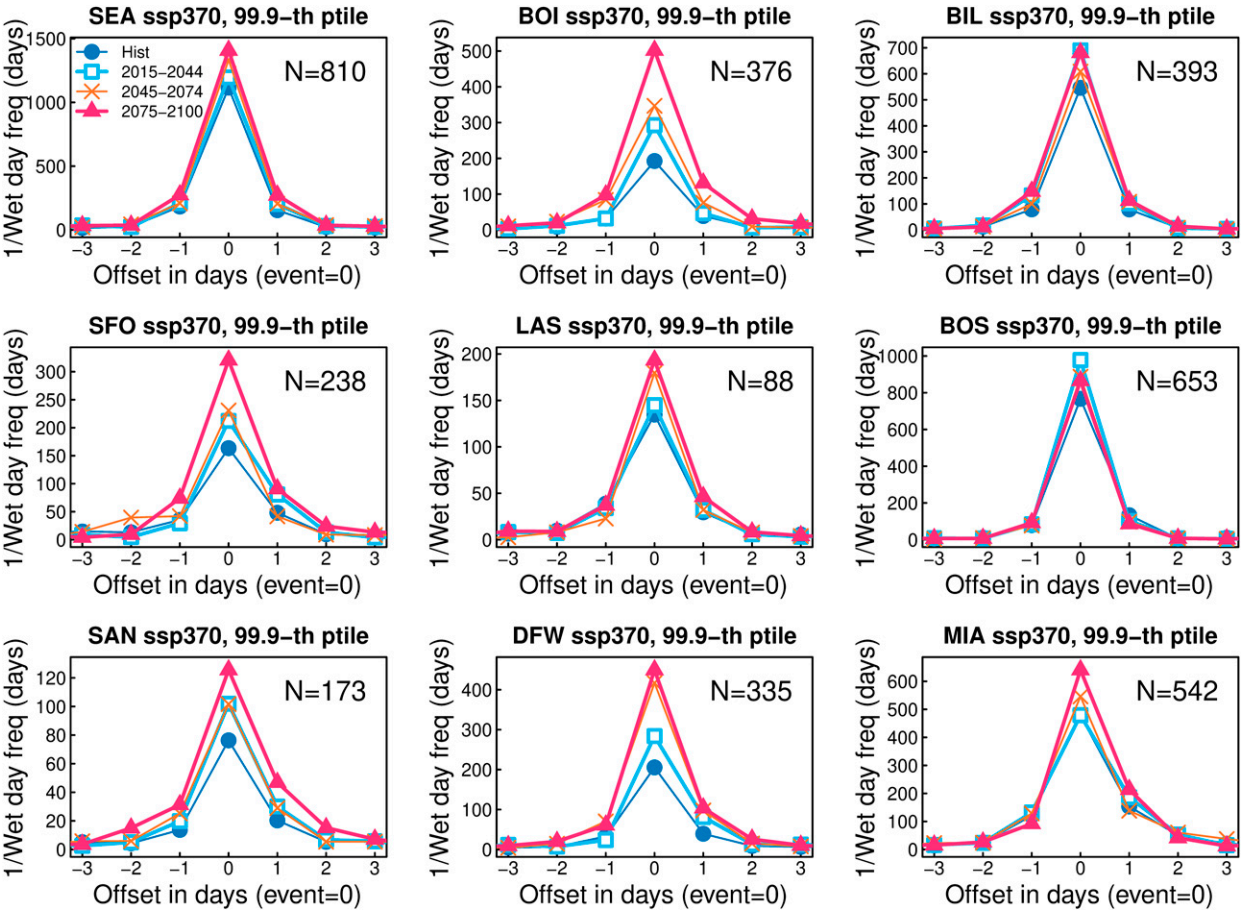

FIG. 13. Multimodel ensemble wet day return period (1/frequency; days) averaged over a box ±30 grid cells (~134 000 km^2) centered on the station indicated in the panel title. Panels are arranged roughly geographically. Values are averaged over all storm events where the center station grid cell experienced at least a 99.9th percentile wet day. The *x* axis shows the time evolution of the area-averaged wet day return period in the three days leading up to the storm, the day of the storm (offset = 0), and the three days after the storm. As per the legend, the dark blue curve is for the historical period, light blue is 2015–44, orange is 2045–74, and the magenta is 2075–2100, all for SSP 370. Although each period has a different number of events, the average number (*N*) across all periods is shown on each panel.

(middle right) fares better, capturing some of the higher coastal values, but still has a mean 18% less than the gridded stations. As discussed in Pierce et al. (2021), not splitting precipitation observations across days ameliorates the reduction in extremes found in L15, but the gridding process still reduces the return values.

The next steps after constructing the training dataset are bias correction and statistical downscaling. The MMEA result from the LOCA version 1 downscaled data (CMIP5; lower left) has a mean 20-yr return value another 9 percentage points lower than the L15 training data used in that work, while LOCA version 2 (CMIP6, lower right) has a mean 6 percentage points less than the Pierce et al. (2021) training data used here.

To summarize, the biggest factor in reducing extremes compared to return values calculated directly from the station data is the gridding process when creating the daily training dataset. Second, not splitting observations across days reduces the error by about a third [29% reduction in L15 versus 18% reduction in Pierce et al. (2021)]. Finally, the ensemble bias correction approach results in less error from the bias correction step, by about one-third compared to the original approach, since in LOCA version 1 the bias correction step reduced extremes by 9 percentage points while in LOCA version 2 the bias correction step reduces extremes by only 6 percentage points. However, the errors introduced by the bias correction are considerably smaller than those arising from the training data gridding process. Overall, updates to the extreme precipitation estimates in LOCA2 suggest that end-users who developed analyses based on the older LOCA-CMIP5 data might find it useful to update their calculations to reflect the higher LOCA2-CMIP6 values.

4. Discussion

The ensemble bias correction approach used here is useful but not a panacea for the problem of bias correcting precipitation extremes in a heterogenous dataset with a variable number

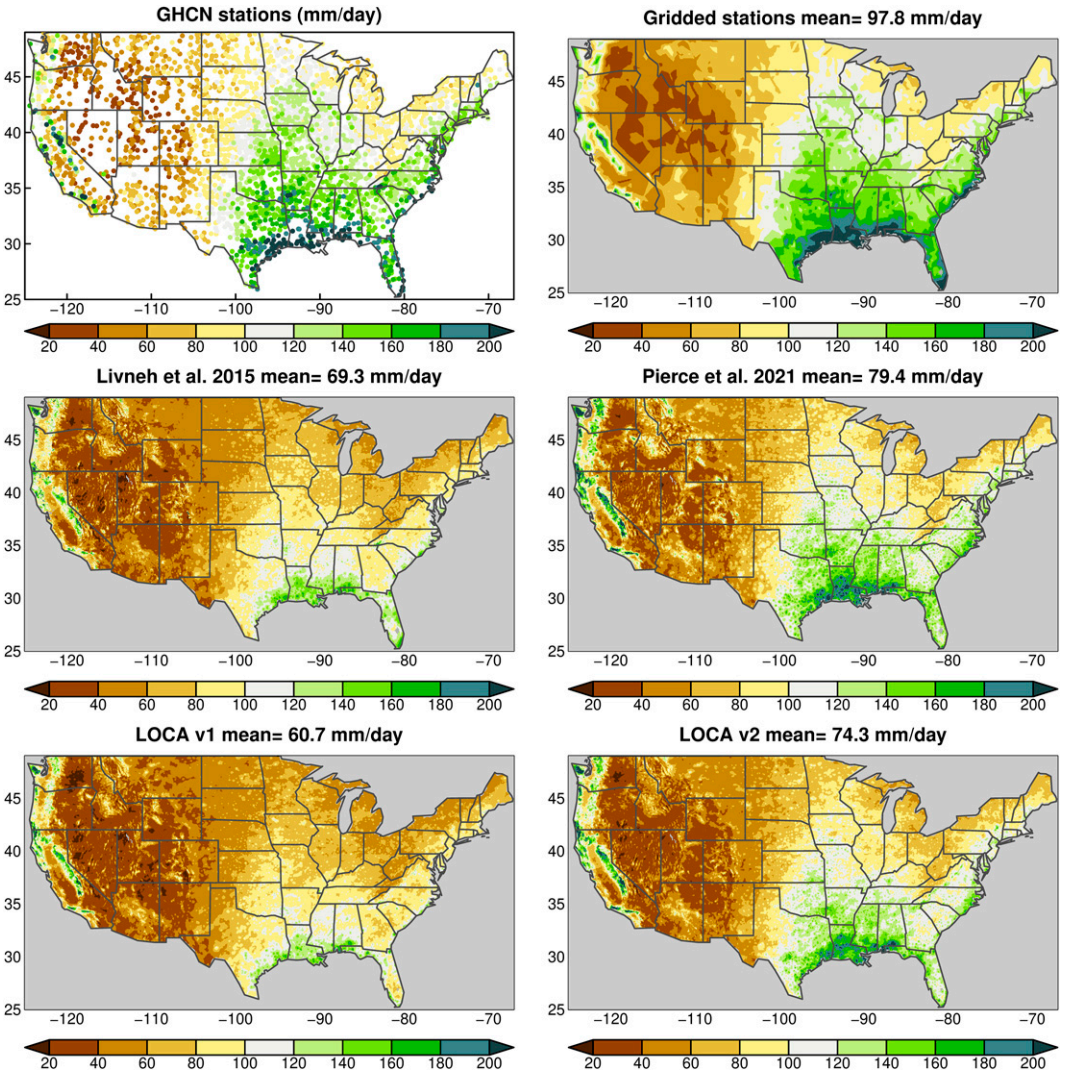

FIG. 14. The 20-yr return value (mm day^{-1}) of daily precipitation over the CONUS for six datasets. (top left) 3662 GHCN stations. (top right) the GHCN station data gridded to the LOCA grid. (middle left) L15. (middle right) Pierce et al. (2021). (bottom left) Multimodel ensemble average (MMEA) 20-yr return value from LOCA version 1 (CMIP5). (bottom right) MMEA from LOCA version 2 (CMIP6).

of ensemble members per model. First, since every model has its own biases (e.g., Fig. 1) it is not obvious how one could combine data from all the models together into one coherent distribution. A model with one ensemble member could have an unusually large extreme value because of biases, or because the model has realistic extremes but the one ensemble member available just happened to have an exceedingly rare high precipitation value. Second, the varying number of ensemble members across models means that the rarest precipitation values are more likely to be found in models with many ensemble members simply because the pool of data is larger. This is not a consequence of the ensemble bias correction approach but rather is a characteristic of the CMIP6 dataset itself, since there is a greater chance of finding rare extreme values in a model that provides 10 ensemble members than in a model that provides only 1 ensemble member. Ensemble bias correction preserves this characteristic, which is realistic but nonetheless different from the results that would have been obtained with a nonensemble bias correction approach. The upshot is that practitioners may want to consider how model selection for an application involving precipitation extremes will be influenced by the number of ensemble members available and the likelihood of finding extreme values in a model that provides more data to analyze.

The overall portrait of extreme precipitation day changes described here exhibits increased extremes over much of North America but with variability across season, location, time horizon, and emissions scenario. This variability must be considered in anticipating how a specific location will be affected by climate change. However, projected increases in extreme day precipitation are substantial over most of the CONUS and Canada, lending credence to the concern that

TABLE 3. Multimodel ensemble average 50-year return value of daily precipitation (mm day^{-1}) for the 16 stations over the historical period, midcentury, and end of century for the SSP 370 emissions scenario. The estimated 95% confidence interval is given in parentheses.

Station	Historical (1950–2014)	2045–74	2075–2100
ATL	104.2 (87.3–125.5)	122.8 (89.9–170.9)	120.0 (90.4–162.2)
BIL	65.5 (48.6–89.7)	74.2 (46.5–118.9)	79.1 (48.8–126.7)
BOI	32.6 (26.5–40.7)	37.2 (26.4–53.8)	38.0 (27.0–55.2)
BOS	131.0 (101.2–172.9)	151.2 (101.7–228.9)	154.5 (104.7–231.1)
CHI	113.3 (88.7–146.3)	128.0 (87.9–188.1)	131.8 (89.1–197.1)
DCA	144.3 (103.4–204.8)	162.4 (102.1–259.8)	166.8 (104.4–269.2)
DEN	65.2 (52.1–81.9)	68.1 (49.5–93.9)	70.5 (50.9–99.4)
DFW	144.4 (111.2–190.5)	162.9 (107.7–249.3)	167.9 (109.2–259.4)
ELP	53.2 (42.8–66.1)	56.0 (39.3–80.6)	56.8 (38.2–85.0)
FAR	79.6 (66.5–95.9)	85.1 (64.0–114.7)	89.6 (65.0–126.1)
LAS	35.7 (29.3–43.5)	37.0 (26.8–51.9)	41.9 (28.9–61.2)
MCI	148.5 (109.8–205.2)	161.3 (102.3–255.1)	160.2 (101.1–255.3)
MIA	179.9 (137.6–237.3)	194.6 (128.9–295.9)	203.6 (126.5–327.5)
SAN	54.0 (47.5–61.1)	57.3 (46.8–70.1)	59.8 (46.6–77.2)
SEA	89.6 (66.6–123.1)	107.0 (72.1–161.5)	110.0 (74.8–163.4)
SFO	76.8 (65.3–91.3)	80.6 (62.5–105.6)	87.8 (65.2–120.2)

these changes will pose a challenge for managing water infrastructure for flood mitigation. Furthermore, the ensemble bias correction methodology used here does a better job at representing those extremes than a nonensemble approach where all maxima are set to the same value, giving practitioners a useful tool when trying to mitigate potential water management problems arising from the increases in extreme day precipitation.

As an example of challenges that might be faced at individual locations, multimodel ensemble average projected 50-yr return values of daily precipitation at the 16 stations illustrated in Fig. 2 are given in Table 3 for the historical period, midcentury, and end of this century under SSP 370. At most locations the projected increases are substantial, for example, an increase from 113 to 132 mm at Chicago, Illinois, or 104 to 120 mm at Atlanta, Georgia, by end of century. Although the general sense of increase is consistent across the stations, there are significant uncertainties associated with these estimates, as shown by the values in the parentheses, which give the 95% confidence interval. Even the midcentury increases under this middle emissions scenario (of the three we considered) are substantial enough to change the probably of flooding at many locations across the CONUS, indicating that existing water management infrastructure designed using historical conditions may not perform well in future decades.

The current version of bias correction outlined here does not take into account nonstationarity over the observed historical period when fitting observed extremes. As climate changes become stronger over time, this will become a progressively less appropriate assumption that should be revisited, for example by fitting the observations to extreme value distributions with time-dependent parameters as in Risser et al. (2021).

5. Conclusions

Future changes in extreme daily precipitation are a key aspect of climate change, with applications to flood preparedness and water management infrastructure. This information is available from numerous models in the CMIP6 archive at coarse resolution, but many practitioners require more localized climate projections and furthermore may rely upon calibrated impact or operations models that require bias-corrected input.

Using the localized constructed analogs (LOCA) statistical downscaling method, we have downscaled data from 27 models, up to 10 ensemble members each, covering SSPs 245, 370, and 585 (329 runs total, where each run is a combination of model/ensemble/SSP). Compared to the version of LOCA used to downscale the CMIP5 models (Pierce et al. 2014), the new effort incorporates ensemble bias correction to better capture daily precipitation extremes and uses an improved daily precipitation training dataset (Pierce et al. 2021) that itself better represents observed precipitation extremes. Together, these improvements give a better representation of daily precipitation extremes in the downscaled data, although our analysis also points to the need for improved methods of gridding daily station observations that preserve the magnitude of rare precipitation events.

The downscaled data show increasing precipitation extremes across much of the CONUS and Canada, although details vary across location, season, and rarity of the precipitation event. In general, 20%–30% increases in winter, spring, and autumn 5- to 500-yr return values of daily precipitation are seen in Canada and the CONUS exclusive of Arizona and New Mexico. Both projected increases and model agreement on the sign of the projected change are greatest in winter (DJF), which will be problematic in locations where flooding is historically worst in winter, particularly since damage increases exponentially as precipitation amounts and runoff increase (Corringham et al. 2019). On the other hand, since many locations exhibit a tendency toward fewer wet days and a drop in the frequency of days with lighter precipitation amounts, the increase in frequency and magnitude of relatively infrequent heavy precipitation days can provide a vital counteracting boost to water supplies. Summer is the most complicated season, with little projected change in 5-yr return

values outside of New England but substantial increases in 500-yr return values across most of the CONUS. In the southern part of the domain, including Mexico, Arizona, and New Mexico, the greatest increases are seen in autumn rather than winter, with values ranging from 10% to 30% depending on return period. Five-year return values decline 15% in summer but show little change at longer return periods.

These results will be useful for climate change planning and mitigation efforts across much of North America, from central Mexico through southern Canada. With its greater number of projections, the new, CMIP6 version of the downscaled data provides useful, more accurate depictions of extreme daily precipitation than the previous LOCA downscaled CMIP5 dataset.

Acknowledgments. This project was funded by a contract from the Strategic Environmental Research and Development Program (SERDP), Project RC19-1391. Additional support for DWP and DRC was provided by the California Energy Commission under Contract EPC-20-006, and from the California Department of Water Resources Atmospheric River Program Phase III under Grant 4600014294. The NASA High-End Computing Capability (HECC) Program provided resources supporting this work through the NASA Earth Exchange (NEX), Earth Science Division, and the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. We thank the NEX staff for their technical support. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. Comments from anonymous reviewers on an initial draft of this work are gratefully acknowledged.

Data availability statement. The daily minimum and maximum temperature and precipitation data are available via Globus at the endpoint https://app.globus.org/file-manager?origin_id=8b53db8a-d0e2-11ec-b95d-0f43df60473d&origin_path=%2F%7E%2Floca2-gdo%2F, and are currently being transferred for public distribution from the Green Data Oasis archive hosted at Lawrence Livermore National Laboratory and operated in conjunction with the U.S. Bureau of Reclamation (https://gdo-dcp.ucllnl.org/downscaled_cmip_projections/dcpInterface.html).

REFERENCES

Abatzoglou, J. T., and T. J. Brown, 2012: A comparison of statistical downscaling methods suited for wildfire applications. *Int. J. Climatol.*, **32**, 772–780, https://doi.org/10.1002/joc.2312.

Ahmed, K. F., G. Wang, J. Silander, A. M. Wilson, J. M. Allen, R. Horton, and R. Anyah, 2013: Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. Northeast. *Global Planet. Change*, **100**, 320–332, https://doi.org/10.1016/j.gloplacha.2012.11.003.

Akinsanola, A. A., G. J. Kooperman, A. G. Pendergrass, W. M. Hannah, and K. A. Reed, 2020: Seasonal representation of extreme precipitation indices over the United States in CMIP6 present-day simulations. *Environ. Res. Lett.*, **15**, 094003, https://doi.org/10.1088/1748-9326/ab92c1.

Arnell, N. W., and S. N. Gosling, 2016: The impacts of climate change on river flood risk at the global scale. *Climatic Change*, **134**, 387–401, https://doi.org/10.1007/s10584-014-1084-5.

Avery, C. W., and Coauthors, 2018: Data tools and scenario products. *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment Volume II*, D. R. Reidmiller et al., Eds., U.S. Global Change Research Program, 1413–1430, https://doi.org/10.7930/NCA4.2018.AP3.

Boé, J. L., L. Terray, F. Habets, and E. Martin, 2007: Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *Int. J. Climatol.*, **27**, 1643–1655, https://doi.org/10.1002/joc.1602.

Bukovsky, M. S., C. M. Carrillo, D. J. Gochis, D. M. Hammerling, R. R. McCrary, and L. O. Mearns, 2015: Toward assessing NARCCAP regional climate model credibility for the North American monsoon: Future climate simulations. *J. Climate*, **28**, 6707–6728, https://doi.org/10.1175/JCLI-D-14-00695.1.

Cannon, A. J., 2018: Multivariate quantile mapping bias correction: An *N*-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dyn.*, **50**, 31–49, https://doi.org/10.1007/s00382-017-3580-6.

Chen, C. A., H.-H. Hsu, and H.-C. Liang, 2021: Evaluation and comparison of CMIP6 and CMIP5 model performance in simulating the seasonal extreme precipitation in the western North Pacific and East Asia. *Wea. Climate Extremes*, **31**, 100303, https://doi.org/10.1016/j.wace.2021.100303.

Chen, J., F. P. Brissette, and R. Leconte, 2014: Assessing regression-based statistical approaches for downscaling precipitation over North America. *Hydrol. Processes*, **28**, 3482–3504, https://doi.org/10.1002/hyp.9889.

——, ——, X. C. Zhang, H. Chen, S. L. Guo, and Y. Zhao, 2019: Bias correcting climate model multi-member ensembles to assess climate change impacts on hydrology. *Climatic Change*, **153**, 361–377, https://doi.org/10.1007/s10584-019-02393-x.

Corringham, T. W., F. M. Ralph, A. Gershunov, D. R. Cayan, and C. A. Talbot, 2019: Atmospheric rivers drive flood damages in the western United States. *Sci. Adv.*, **5**, eaax4631, https://doi.org/10.1126/sciadv.aax4631.

Deser, C., A. Phillips, V. Bourdette, and H. Y. Teng, 2012: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, https://doi.org/10.1007/s00382-010-0977-x.

Dibike, Y. B., and P. Coulibaly, 2005: Hydrologic impact of climate change in the Saguenay watershed: Comparison of downscaling methods and hydrologic models. *J. Hydrol.*, **307**, 145–163, https://doi.org/10.1016/j.jhydrol.2004.10.012.

Dike, V. N., Z. H. Lin, K. Fei, G. S. Langendijk, and D. Nath, 2022: Evaluation and multimodel projection of seasonal precipitation extremes over Central Asia based on CMIP6 simulations. *Int. J. Climatol.*, **42**, 7228–7251, https://doi.org/10.1002/joc.7641.

Donat, M. G., A. L. Lowry, L. V. Alexander, P. A. O'Gorman, and N. Maher, 2016: More extreme precipitation in the world's dry and wet regions. *Nat. Climate Change*, **6**, 508–513, https://doi.org/10.1038/nclimate2941.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Faye, A., and A. A. Akinsanola, 2022: Evaluation of extreme precipitation indices over West Africa in CMIP6 models. *Climate Dyn.*, **58**, 925–939, https://doi.org/10.1007/s00382-021-05942-2.

Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.*, **27**, 1547–1578, https://doi.org/10.1002/joc.1556.

Fritsch, F. N., and R. E. Carlson, 1980: Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.*, **17**, 238–246, https://doi.org/10.1137/0717021.

Ge, F., S. P. Zhu, H. L. Luo, X. F. Zhi, and H. Wang, 2021: Future changes in precipitation extremes over Southeast Asia: Insights from CMIP6 multi-model ensemble. *Environ. Res. Lett.*, **16**, 024013, https://doi.org/10.1088/1748-9326/abd7ad.

Gershunov, A., and Coauthors, 2019: Precipitation regime change in western North America: The role of atmospheric rivers. *Sci. Rep.*, **9**, 9944, https://doi.org/10.1038/s41598-019-46169-w.

Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, 2012: Technical note: Downscaling RCM precipitation to the station scale using statistical transformations – A comparison of methods. *Hydrol. Earth Syst. Sci.*, **16**, 3383–3390, https://doi.org/10.5194/hess-16-3383-2012.

Guo, Q., J. Chen, X. C. Zhang, M. X. Shen, H. Chen, and S. L. Guo, 2019: A new two-stage multivariate quantile mapping method for bias correcting climate model outputs. *Climate Dyn.*, **53**, 3603–3623, https://doi.org/10.1007/s00382-019-04729-w.

Gupta, V., V. Singh, and M. K. Jain, 2020: Assessment of precipitation extremes in India during the 21st century under SSP1-1.9 mitigation scenarios of CMIP6 GCMs. *J. Hydrol.*, **590**, 125422, https://doi.org/10.1016/j.jhydrol.2020.125422.

Hagemann, S., C. Chen, J. O. Haerter, J. Heinke, D. Gerten, and C. Piani, 2011: Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. *J. Hydrometeor.*, **12**, 556–578, https://doi.org/10.1175/2011JHM1336.1.

Harp, R. D., and D. E. Horton, 2022: Observed changes in daily precipitation intensity in the United States. *Geophys. Res. Lett.*, **49**, e2022GL099955, https://doi.org/10.1029/2022GL099955.

Hosking, J. R. M., 1990: L-moment: Analysis and estimation of distributions using linear-combinations of order-statistics. *J. Roy. Stat. Soc.*, **52**, 105–124, https://doi.org/10.1111/j.2517-6161.1990.tb01775.x.

Jacob, D., and Coauthors, 2020: Regional climate downscaling over Europe: Perspectives from the EURO-CORDEX community. *Reg. Environ. Change*, **20**, 51, https://doi.org/10.1007/s10113-020-01606-9.

Kim, Y.-H., S.-K. Min, X. B. Zhang, J. Sillmann, and M. Sandstad, 2020: Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. *Wea. Climate Extremes*, **29**, 100269, https://doi.org/10.1016/j.wace.2020.100269.

Kröner, N., S. Kotlarski, E. Fischer, D. Lüthi, E. Zubler, and C. Schär, 2017: Separating climate change signals into thermodynamic, lapse-rate and circulation effects: Theory and application to the European summer climate. *Climate Dyn.*, **48**, 3425–3440, https://doi.org/10.1007/s00382-016-3276-3.

Kyselý, J., 2008: A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models. *J. Appl. Meteor. Climatol.*, **47**, 3236–3251, https://doi.org/10.1175/2008JAMC1763.1.

Li, F. Y., W. D. Collins, M. F. Wehner, D. L. Williamson, J. G. Olson, and C. Algieri, 2011: Impact of horizontal resolution on simulation of precipitation extremes in an aqua-planet version of Community Atmospheric Model (CAM3). *Tellus*, **63A**, 884–892, https://doi.org/10.1111/j.1600-0870.2011.00544.x.

Li, H. B., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res.*, **115**, D10101, https://doi.org/10.1029/2009JD012882.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land-surface water and energy fluxes for general-circulation models. *J. Geophys. Res.*, **99**, 14 415–14 428, https://doi.org/10.1029/94JD00483.

Livneh, B., T. J. Bohn, D. W. Pierce, F. Munoz-Arriola, B. Nijssen, R. Vose, D. R. Cayan, and L. Brekke, 2015: A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and southern Canada 1950–2013. *Sci. Data*, **2**, 150042, https://doi.org/10.1038/sdata.2015.42.

Luo, N., Y. Guo, J. M. Chou, and Z. B. Gao, 2022: Added value of CMIP6 models over CMIP5 models in simulating the climatological precipitation extremes in China. *Int. J. Climatol.*, **42**, 1148–1164, https://doi.org/10.1002/joc.7294.

Maraun, D., 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *J. Climate*, **26**, 2137–2143, https://doi.org/10.1175/JCLI-D-12-00821.1.

——, 2016: Bias correcting climate change simulations – A critical review. *Curr. Climate Change Rep.*, **2**, 211–220, https://doi.org/10.1007/s40641-016-0050-x.

——, and Coauthors, 2017: Towards process-informed bias correction of climate change simulations. *Nat. Climate Change*, **7**, 764–773, https://doi.org/10.1038/nclimate3418.

Martinkova, M., and J. Kysely, 2020: Overview of observed Clausius-Clapeyron scaling of extreme precipitation in midlatitudes. *Atmosphere*, **11**, 786, https://doi.org/10.3390/atmos11080786.

Maurer, E. P., and D. W. Pierce, 2014: Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrol. Earth Syst. Sci.*, **18**, 915–925, https://doi.org/10.5194/hess-18-915-2014.

——, H. G. Hidalgo, T. Das, M. D. Dettinger, and D. R. Cayan, 2010: The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrol. Earth Syst. Sci.*, **14**, 1125–1138, https://doi.org/10.5194/hess-14-1125-2010.

Mearns, L. O., W. J. Gutowski, R. Jones, L.-Y. Leung, S. McGinnis, A. M. B. Nunes, and Y. Qian, 2009: A regional climate change assessment program for North America. *Eos, Trans. Amer. Geophys. Union*, **90**, 311, https://doi.org/10.1029/2009EO360002.

Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, https://doi.org/10.1175/JTECH-D-11-00103.1.

Michaelis, A. C., A. Gershunov, A. Weyant, M. A. Fish, T. Shulgina, and F. M. Ralph, 2022: Atmospheric river precipitation enhanced by climate change: A case study of the storm that contributed to California's Oroville Dam crisis. *Earth's Future*, **10**, e2021EF002537, https://doi.org/10.1029/2021EF002537.

Michelangeli, P.-A., M. Vrac, and H. Loukos, 2009: Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophys. Res. Lett.*, **36**, L11708, https://doi.org/10.1029/2009GL038401.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, 2008: Stationarity is dead: Whither water management? *Science*, **319**, 573–574, https://doi.org/10.1126/science.1151915.

NCEI, 2022: U.S. Billion-dollar weather and climate disasters, 1980–present (NCEI Accession 0209268). Accessed 9 September 2022, https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:0209268.

Norris, J., G. Chen, and J. D. Neelin, 2019: Thermodynamic versus dynamic controls on extreme precipitation in a warming climate from the Community Earth System Model Large Ensemble. *J. Climate*, **32**, 1025–1045, https://doi.org/10.1175/JCLI-D-18-0302.1.

NRC, 1982: *Scientific Basis of Water-Resource Management*. National Academies Press, 130 pp., https://doi.org/10.17226/19530.

O'Brien, T. A., W. D. Collins, K. Kashinath, O. Rubel, S. Byna, J. M. Gu, H. Krishnan, and P. A. Ullrich, 2016: Resolution dependence of precipitation statistical fidelity in hindcast simulations. *J. Adv. Model. Earth Syst.*, **8**, 976–990, https://doi.org/10.1002/2016MS000671.

O'Neill, B. C., E. Kriegler, K. Riahi, K. L. Ebi, S. Hallegatte, T. R. Carter, R. Mathur, and D. P. van Vuuren, 2014: A new scenario framework for climate change research: The concept of shared socioeconomic pathways. *Climatic Change*, **122**, 387–400, https://doi.org/10.1007/s10584-013-0905-2.

——, and Coauthors, 2016: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.*, **9**, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016.

Ozturk, T., D. Matte, and J. H. Christensen, 2022: Robustness of future atmospheric circulation changes over the EURO-CORDEX domain. *Climate Dyn.*, **59**, 1799–1814, https://doi.org/10.1007/s00382-021-06069-0.

Panofsky, H. A., and G. W. Brier, 1968: *Some Applications of Statistics to Meteorology*. The Pennsylvania State University, 224 pp.

Pierce, D. W., and D. R. Cayan, 2016: Downscaling humidity with localized constructed analogs (LOCA) over the conterminous United States. *Climate Dyn.*, **47**, 411–431, https://doi.org/10.1007/s00382-015-2845-1.

——, T. P. Barnett, B. D. Santer, and P. J. Gleckler, 2009: Selecting global climate models for regional climate change studies. *Proc. Natl. Acad. Sci. USA*, **106**, 8441–8446, https://doi.org/10.1073/pnas.0900094106.

——, and Coauthors, 2013a: The key role of heavy precipitation events in climate model disagreements of future annual precipitation changes in California. *J. Climate*, **26**, 5879–5896, https://doi.org/10.1175/JCLI-D-12-00766.1.

——, and Coauthors, 2013b: Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling. *Climate Dyn.*, **40**, 839–856, https://doi.org/10.1007/s00382-012-1337-9.

——, D. R. Cayan, and B. L. Thrasher, 2014: Statistical downscaling using localized constructed analogs (LOCA). *J. Hydrometeor.*, **15**, 2558–2585, https://doi.org/10.1175/JHM-D-14-0082.1.

——, ——, E. P. Maurer, J. T. Abatzoglou, and K. C. Hegewisch, 2015: Improved bias correction techniques for hydrological simulations of climate change. *J. Hydrometeor.*, **16**, 2421–2442, https://doi.org/10.1175/JHM-D-14-0236.1.

——, J. F. Kalansky, and D. R. Cayan, 2018: Climate, drought, and sea level rise scenarios for California's fourth climate change assessment. California Energy Commission Rep. CCCA4-CEC-2018-006, 78 pp., https://www.energy.ca.gov/sites/default/files/2019-11/Projections_CCCA4-CEC-2018-006_ADA.pdf.

——, L. Su, D. R. Cayan, M. D. Risser, B. Livneh, and D. P. Lettenmaier, 2021: An extreme-preserving long-term gridded daily precipitation dataset for the conterminous United States. *J. Hydrometeor.*, **22**, 1883–1895, https://doi.org/10.1175/JHM-D-20-0212.1.

——, D. R. Cayan, J. Goodrich, T. Das, and A. Munevar, 2022: Evaluating global climate models for hydrological studies of the upper Colorado River Basin. *J. Amer. Water Resour. Assoc.*, **58**, 709–734, https://doi.org/10.1111/1752-1688.12974.

Polade, S. D., A. Gershunov, D. R. Cayan, M. D. Dettinger, and D. W. Pierce, 2017: Precipitation in a warming world: Assessing projected hydro-climate changes in California and other Mediterranean climate regions. *Sci. Rep.*, **7**, 10783, https://doi.org/10.1038/s41598-017-11285-y.

Ralph, F. M., and Coauthors, 2014: A vision for future observations for western U.S. extreme precipitation and flooding. *J. Contemp. Water Res. Educ.*, **153**, 16–32, https://doi.org/10.1111/j.1936-704X.2014.03176.x.

Riahi, K., and Coauthors, 2017: The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environ. Change*, **42**, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009.

Risser, M. D., D. R. Feldman, M. F. Wehner, D. W. Pierce, and J. R. Arnold, 2021: Identifying and correcting biases in localized downscaling estimates of daily precipitation return values. *Climatic Change*, **169**, 33, https://doi.org/10.1007/s10584-021-03265-z.

Sachindra, D. A., K. Ahmed, M. M. Rashid, S. Shahid, and B. J. C. Perera, 2018: Statistical downscaling of precipitation using machine learning techniques. *Atmos. Res.*, **212**, 240–258, https://doi.org/10.1016/j.atmosres.2018.05.022.

Screen, J. A., T. J. Bracegirdle, and I. Simmonds, 2018: Polar climate change as manifest in atmospheric circulation. *Curr. Climate Change Rep.*, **4**, 383–395, https://doi.org/10.1007/s40641-018-0111-4.

Sharifi, E., B. Saghafian, and R. Steinacker, 2019: Downscaling satellite precipitation estimates with multiple linear regression, artificial neural networks, and spline interpolation techniques. *J. Geophys. Res. Atmos.*, **124**, 789–805, https://doi.org/10.1029/2018JD028795.

Srivastava, A., R. Grotjahn, and P. A. Ullrich, 2020: Evaluation of historical CMIP6 model simulations of extreme precipitation over contiguous US regions. *Wea. Climate Extremes*, **29**, 100268, https://doi.org/10.1016/j.wace.2020.100268.

Srivastava, A. K., R. Grotjahn, P. A. Ullrich, and M. Sadegh, 2021: Pooling data improves multimodel IDF estimates over median-based IDF estimates: Analysis over the Susquehanna and Florida. *J. Hydrometeor.*, **22**, 971–995, https://doi.org/10.1175/JHM-D-20-0180.1.

Sturman, A., and H. Quenol, 2013: Changes in atmospheric circulation and temperature trends in major vineyard regions of New Zealand. *Int. J. Climatol.*, **33**, 2609–2621, https://doi.org/10.1002/joc.3608.

Tan, Y. H., F. Zwiers, S. Yang, C. Li, and K. Q. Deng, 2020: The role of circulation and its changes in present and future atmospheric rivers over western North America. *J. Climate*, **33**, 1261–1281, https://doi.org/10.1175/JCLI-D-19-0134.1.

Teutschbein, C., and J. Seibert, 2012: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *J. Hydrol.*, **456–457**, 12–29, https://doi.org/10.1016/j.jhydrol.2012.05.052.

Thrasher, B., E. P. Maurer, C. McKellar, and P. B. Duffy, 2012: Technical note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrol. Earth Syst. Sci.*, **16**, 3309–3314, https://doi.org/10.5194/hess-16-3309-2012.

U.S. Bureau of Reclamation, 2013: Downscaled CMIP3 and CMIP5 climate and hydrology projections: Release of downscaled CMIP5 climate projections, comparison with preceding information, and summary of user needs. USBR Tech. Service Center, 47 pp., https://gdo-dcp.ucllnl.org/downscaled_cmip_projections/techmemo/LOCA_BCSD_hydrology_tech_memo.pdf.

Vaittinada Ayar, P., M. Vrac, and A. Mailhot, 2021: Ensemble bias correction of climate simulations: Preserving internal variability. *Sci. Rep.*, **11**, 3098, https://doi.org/10.1038/s41598-021-82715-1.

Van den Dool, H. M., 1994: Searching for analogs, how long must we wait. *Tellus*, **46A**, 314–324, https://doi.org/10.1034/j.1600-0870.1994.t01-2-00006.x.

Van Uytven, E., J. D. Niel, and P. Willems, 2020: Uncovering shortcomings of a weather typing method. *Hydrol. Earth Syst. Sci.*, **24**, 2671–2686, https://doi.org/10.5194/hess-24-2671-2020.

Vano, J. A., M. D. Dettinger, R. Cifelli, D. Curtis, A. Dufour, K. Miller, J. R. Olsen, and A. M. Wilson, 2019: Hydroclimatic extremes as challenges for the water management community: Lessons from Oroville Dam and Hurricane Harvey. *Bull. Amer. Meteor. Soc.*, **100**, S9–S14, https://doi.org/10.1175/BAMS-D-18-0219.1.

——, and Coauthors, 2020: Comparing downscaled LOCA and BCSD CMIP5 climate and hydrology projections. U.S. Bureau of Reclamation, 96 pp., https://gdo-dcp.ucllnl.org/downscaled_cmip_projections/techmemo/LOCA_BCSD_hydrology_tech_memo.pdf.

Wang, G., C. J. Kirchoff, A. Seth, J. T. Abatzoglou, B. Livneh, D. W. Pierce, L. Fomenko, and T. Ding, 2020: Projected changes of precipitation characteristics depend on downscaling method and training data: MACA versus LOCA using the U.S. Northeast as an example. *J. Hydrometeor.*, **21**, 2739–2758, https://doi.org/10.1175/JHM-D-19-0275.1.

Westra, S., and Coauthors, 2014: Future changes to the intensity and frequency of short-duration extreme rainfall. *Rev. Geophys.*, **52**, 522–555, https://doi.org/10.1002/2014RG000464.

Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.*, **107**, 4429, https://doi.org/10.1029/2001JD000659.

——, L. R. Leung, V. Sridhar, and D. P. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, **62**, 189–216, https://doi.org/10.1023/B:CLIM.0000013685.99609.9e.

Xu, H. W., H. P. Chen, and H. J. Wang, 2022: Future changes in precipitation extremes across China based on CMIP6 models. *Int. J. Climatol.*, **42**, 635–651, https://doi.org/10.1002/joc.7264.

Yang, Y., and Coauthors, 2021: Atmospheric circulation patterns conducive to severe haze in eastern China have shifted under climate change. *Geophys. Res. Lett.*, **48**, e2021GL095011, https://doi.org/10.1029/2021GL095011.

Yin, J. H., 2005: A consistent poleward shift of the storm tracks in simulations of 21st century climate. *Geophys. Res. Lett.*, **32**, L18701, https://doi.org/10.1029/2005GL023684.

Zappa, G., 2019: Regional climate impacts of future changes in the mid-latitude atmospheric circulation: A storyline view. *Curr. Climate Change Rep.*, **5**, 358–371, https://doi.org/10.1007/s40641-019-00146-7.

Zhu, H. H., Z. H. Jiang, J. Li, W. Li, C. X. Sun, and L. Li, 2020: Does CMIP6 inspire more confidence in simulating climate extremes over China? *Adv. Atmos. Sci.*, **37**, 1119–1132, https://doi.org/10.1007/s00376-020-9289-1.